Threshold Autoregressive Model with Multiple Threshold Variables

CHEN Haiqiang

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Master of Philosophy
in
Economics

July 2005

ABSTRACT

This thesis considers issues related to the estimation and inference for a threshold autoregressive model with multiple threshold variables. Regime shifts can occur among all of the categories split by the threshold variables. Certain restrictions on structural parameters are also allowed so that changes just occur among a subset of regimes. It is shown that, under some regularity conditions, the OLS threshold estimators are consistent. We also derive the limiting joint distribution of the threshold estimators, using an early result in a two-regime threshold model of Chong and Yan (2004). Of special interest is the fact that the estimators are consistent even there are misspecifications in both the number of regimes and the functional form of the regression equation. Based on this result, we present a sequential model-building procedure. A likelihood ratio test is also proposed to detect the threshold effect and determine the number of regimes. Monte Carlo simulations evaluate the performance of the above estimators and the test in the finite-sample case. An application to the stock market is provided. The result suggests the existence of a possible four-regime cycle in the return series of stocks.

摘要

本論文考慮了多元閾值自回歸模型的估計和推斷問題。 在我們的模型中，區域跳躍可以發生在所有區域之間，但是對模型參數加入一些約束條件後，跳躍也可以只發生在某些特定區域之間。 在一些常見的假設下，我們證明瞭閾估計變數的一致性，同時，我們也得到了其漸近分佈，擴展了 Chong & Yan (2004)的結果。最有趣的是,即使估計模型使用了錯誤的區域數目和回歸變數的方程形式, 閾估計變數仍然是一致的。 在這個穩健的估計結果基礎上, 我們提出了一個逐步建模的方法。 爲了檢驗模型中有沒有區域跳躍，我們設計了一個似然比檢驗。 所有的理論結果都用蒙特卡羅實驗得到了驗證。 我們用這個模型去擬合股市指數，結果表明股市指數中可能存在一個四區域週期。

ACKNOWLEDGEMENTS

I would like to express my honest gratitude to my supervisor, Professor Chong Tai-Leung for his invaluable guidance, discussions and support through every stage of my thesis. I would also like to pay special thanks to Professor Chou Win-Lin and Professor Zeng Zhixiong for their important comments and suggestions that have improved the quality of my work. Meanwhile, I wish to thank my fellow classmates for their kind help and encouragement.

CONTENTS

1 Introduction

Threshold autoregressive model has received much attention during past decades. It is first proposed by Tong (1978) and discussed by Tong and Lim (1980) and Tong (1983) in detail. Many models are designed in order to deal with various types of data in reality, such as the smooth transition threshold autoregressive model (STAR) of Chan and Tong (1986), the exponential autoregressive (EXPAR) model of Haggan and Ozaki (1981) and the functional-coefficient autoregressive (FAR) model of Chen and Tsay (1993). Threshold autoregressive model has a wide range of applications. For example, Henry, Olekaln and Summers (2001) provide evidence of threshold nonlinearity in the Australian real exchange rate. Potter (1995) uses a two-regime threshold autoregressive model to analyze the fluctuations of the U.S. economic output. The multivariate threshold autoregressive models are designed by Tsay (1998) to study the arbitrage activities of security market. More recently, Dueker, Sola and Spagnolo (2003) propose a contemporaneous TAR model and apply it to the pricing of bonds.

Most of the above models are based on a single threshold variable only. Few studies, however, have been carried out to explore models with multiple threshold variables. For a single threshold model, Tsay (1998) argues that it can be transformed into a standard change-point model by re-indexing the threshold variable. His approach, however, does not work when there are more than one threshold variable. Moreover, if threshold variables are dependent or if they are correlated with the regressors, the statistical properties of the estimators are more complicated than those of the single threshold model. As a result, it is difficult to extend the result of the single-threshold variable case to the multiple one.

Chong and Yan (2004) first investigate the statistical properties of the threshold estimators in a regression model with multiple threshold variables and employ the model to predict financial crises. Their work provides a theoretical foundation for this kind of model. However, the model is still simple, with only two regimes considered, and it also requires that the specific form of the model be known before the estimation, which is not typical in reality. In an time series threshold model, the number of regimes and the order are both unknown.[1] Tsay (1998) uses a grid search method to locate the threshold variable and select the order of the model simultaneously by minimizing the sum of the Akaike Information Criterion (AIC) of each regime. The method needs to assume the number of regimes known. Meanwhile, it could not obtain consistent threshold estimators and the computational workload is very huge when there are more than one threshold variable.

This thesis considers these issues by proposing a new kind of threshold autoregressive model with multiple threshold variables, which is defined specifically in Section 2. Here we denote the model as M-TAR[2] model, where M stands for the multiple threshold variables. In this model, economic time series are allowed to undergo regime shifts based on the values of more than two threshold variables. An obvious example is that x_t is a process with several regime shifts depending on the values of x_{t-1} and x_{t-2}. We set the number of threshold variables to two but leave the number of regimes unknown.[3] Two threshold variables can split the sample into four categories at most. We refer to the threshold model as the full model when shifts occur among all of the categories. Certain restrictions on structural parameters are allowed so that changes just occur in a subset of regimes, and then the model becomes a restricted model, with just two regimes or three regimes. This work takes a four-regime model as the basic model and related results can be extended to restricted models without much difficulty.

It is shown that, under some regular conditions, the least squares threshold estimators are strongly consistent. With a similar approach used by Hansen (2000)

[1] m threshold variables can split the sample into 2^m regimes at most . The number of regimes is usually unknown when there are multiple threshold variables.

[2] Enders and Granger (1998) refer to the momentum threshold model as MTAR model.

[3] Chong and Yan (2004)'s model lets the number of regimes be two and leave the number of threshold variables unknown. Our model assumes that there are only two threshold variables, but it can be extended to the case with more than two threshold variables. The procedure is not mathematically difficult but very tedious. Meanwhile, the tests developed by Chong and Yan (2004) could be applied here to select threshold variables.

and Chong and Yan (2004), we have derived the asymptotic approximation to the joint distribution of the least-squares estimator $\hat{\gamma}$ of the threshold parameter vector γ. The distribution has a similar form as that of Chong and Yan (2004) but with a different scale.[4]

We show that the estimators are consistent even threshold models are misspecified in both the number of regimes and the functional form of regression equation. Based on this result, we present a sequential model-building procedure. Compared with the method estimating all of parameters simultaneously, the sequential procedure could save much workload. We also propose likelihood ratio type statistics to test the null hypothesis of no threshold effect and to determine the number of regimes. Monte Carlo simulations are presented to highlight the performance of the modeling procedure and the test in the finite-sample case.

An empirical application of the model on the stock market index is given. It has long been observed that the market index may experience several regime shifts. An obvious example is that investors usually classify the market as the bull market or the bear market according to the past performance of the market index. A lot of researches have shown that stocks behave differently in these two types of market. For example, Kim and Zumwalt (1979) show that the betas that measure the risk of the securities are different in the bull and the bear market. Lunde and Timmermann (2004) study the long-run serial correlations of stock returns and show that duration dependence exists in return series. Threshold models have been used to describe these asymmetry between the bull and the bear market. Related studies include the threshold autoregressive conditional heteroscedastic (TARCH) model of Li and Lam (1995), the threshold error-correction model of Markens et al. (1998) and the threshold random walk model of Shively (2003).

We apply the M-TAR model to Hang Seng Index (HSI).[5] The model incorpo-

[4]Chong and Yan (2004) 's model is a restricted model when there are only two threshold variables. The papremeter restriction affects the scalé in asymptotic distribution, but it would not produce any effect on the consistency of the estimators.

[5]The stock index of Hong Kong.

rates more information from market than previous studies by using past turnover as one of the threshold variables. Turnover is another key indicator, besides price, to describe the performance of the market and it plays an important role in classifying the market.[6] Yet, no empirical result from threshold models reflects this role of the turnover. Our model fills this gap and detects a threshold effect in the return series of HSI. The result suggests a four-regime cycle that produces a reasonable explanation to the nonlinearity found in the return series of stocks.

The rest of the thesis is organized as follows: Section 2 presents the basic model with some regular assumptions. Section 3 examines the consistency of the threshold estimators in our model. Section 4 studies the asymptotic distribution of the estimators and related test for the threshold. Section 5 provides a modeling procedure and discusses the consistency of the estimators in the presence of model misspecifications. Section 6 gives Monte Carlo simulation experiments. Section 7 presents an application of the model. Section 8 concludes. All proofs are relegated to the Appendix.

2 The Model

Assume x_t satisfies the following threshold autoregressive model with two threshold variables and four regimes:

$$x_t = \sum_{j=1}^{4} \Psi_t^{(j)} \left(\gamma^0\right) (\phi_0^{(j)} + \sum_{i=1}^{p} \phi_i^{(j)} x_{t-i}) + u_t, \tag{1}$$

where

$$\begin{aligned}
\Psi_t^{(1)}\left(\gamma^0\right) &= I\left(z_{1t} \leq \gamma_1^0, z_{2t} \leq \gamma_2^0\right); \\
\Psi_t^{(2)}\left(\gamma^0\right) &= I\left(z_{1t} \leq \gamma_1^0, z_{2t} > \gamma_2^0\right); \\
\Psi_t^{(3)}\left(\gamma^0\right) &= I\left(z_{1t} > \gamma_1^0, z_{2t} \leq \gamma_2^0\right); \\
\Psi_t^{(4)}\left(\gamma^0\right) &= I\left(z_{1t} > \gamma_1^0, z_{2t} > \gamma_2^0\right).
\end{aligned}$$

Here, we assume $p_1 = p_2 = p_3 = p_4 = p$, where p_i is the order in each regime, but this assumption can be relaxed easily by setting $p = \max(p_1, p_2, p_3, p_4)$, and $\phi_q^{(j)} = 0$ when $q > p_j$, $j = 1, 2, 3, 4$.

[6]See Karpoff (1987), Lee and Swaminathan (2000).

$\gamma^0 = \left(\gamma_1^0, \gamma_2^0\right)$ is the threshold parameter vector pending to be estimated; $z_t = (z_{1t}, z_{2t})$ are the threshold variables; $\phi^{(j)} = (\phi_1^{(j)}, \phi_2^{(j)}, ..., \phi_p^{(j)})$ are the structural parameters and $\phi^{(i)} \neq \phi^{(j)}$ when $i \neq j$. Note that some items in $\phi^{(j)}$ may be zero. $\Psi_t^{(j)}(\gamma^0)$ is the threshold condition, which equals one when some requirements are satisfied, and equals zero otherwise. Such a process partitions one-dimension Euclidean space into 2×2 regimes. In each regime, the model is an AR model. Since the two-dimension threshold space cannot be sorted by one variable, the model cannot be converted to a standard structural-change problem through the arranged regression approach proposed by Tsay (1998).

Model (1) is a very general setting. The threshold variables z_1 and z_2 could be variables external to the system or lagged values of x_t, for example, $z_1 = x_{t-d_1}$, and $z_2 = x_{t-d_2}$.[7] Thus, the model includes the self-exciting TAR (SETAR) model proposed by Tong (1983). When $z_1 = z_2$, the model becomes the multivariate threshold model proposed by Tsay (1998). We can also allow certain restrictions on structural parameters so that the changes just occur in a subset of regimes, and then the model becomes a restricted model with two or three regimes. For example, set $\phi^{(1)} = \phi^{(2)} = \phi^{(3)}$, the model will be simplified as a threshold model with two regimes. This is the case studied by Chong and Yan (2004) when there are only two threshold variables.

The model could be rewritten as

$$x_t = \left\{ \begin{array}{l} \phi_0^{(1)} + \phi_1^{(1)} x_{t-1} + \phi_2^{(1)} x_{t-2}, ..., +\phi_p^{(1)} x_{t-p} + u_t, \text{when } z_{1t} \leq \gamma_1^0, z_{2t} \leq \gamma_2^0 \\ \phi_0^{(2)} + \phi_1^{(2)} x_{t-1} + \phi_2^{(2)} x_{t-2}, ..., +\phi_p^{(2)} x_{t-p} + u_t, \text{ when } z_{1t} \leq \gamma_1^0, z_{2t} > \gamma_2^0 \\ \phi_0^{(3)} + \phi_1^{(3)} x_{t-1} + \phi_2^{(3)} x_{t-2}, ..., +\phi_p^{(3)} x_{t-p} + u_t, \text{ when } z_{1t} > \gamma_1^0, z_{2t} \leq \gamma_2^0 \\ \phi_0^{(4)} + \phi_1^{(4)} x_{t-1} + \phi_2^{(4)} x_{t-2}, ..., +\phi_p^{(4)} x_{t-p} + u_t, \text{ when } z_{1t} > \gamma_1^0, z_{2t} > \gamma_2^0 \end{array} \right\}.$$

Given the observation data of (x_t, z_t), $t = 1, 2, ..., T$, our objective is to estimate the threshold parameters γ^0 and the structural parameters $\phi^{(j)} = (\phi_1^{(j)}, \phi_2^{(j)}, ..., \phi_p^{(j)})$. The model is linear within a regime but nonlinear across regimes. It can be estimated in each regime using least squares method conditional on the threshold variables and the number of regimes.

[7] d_1 and d_2 are positive integers referred to as the delay or threshold lag.

Before proceeding to the estimation, we impose the following assumptions to ensure the consistency of estimators and the identification of unknown variables.

$(A1)$ All roots of the characteristic equation $y^p - \phi_1^{(j)} y^{p-1} - \phi_2^{(j)} y^{p-2}, ..., -\phi_p^{(j)} = 0$, $j = 1, 2, 3, 4$, lie strictly inside the unit circle;

$(A2)$ u_t is a white noise sequence with zero means and variance σ^2, and $E\left|u_t\right|^4 < \infty$ for all t;

$(A3)$ $\frac{1}{T}\sum_{t=1}^{T}(x_t - \overline{x})^2 \xrightarrow{p} \sigma_x^2$ and $\frac{1}{T}\sum_{t=[aT]+1}^{[bT]}(x_t - \overline{x})^2 \xrightarrow{p} (b-a)\sigma_x^2$, $0 < a < b < 1$;

$(A4)$ Denote n_j as the number of the observations that are in the jth regime. Assume that

$$\frac{n_j}{T} \rightarrow c_j \text{ in probability, for all } j = 1, 2, 3, 4. \tag{2}$$

where c_j is a non-negative fraction and $\sum_{j=1}^{4} c_j = 1$.

Assumptions $(A1)$ to $(A3)$ ensures that x_t be stationary in each regime and provide the moment condition for uniform convergence results. $(A4)$ states that observations fall into each regime with a certain probability.

3 Least Squares Estimation

In this section, we show the consistency of least squares estimators. Let

$$x_t = \sum_{j=1}^{4} \Psi_t^{(j)}\left(\gamma^0\right)(\phi_0^{(j)} + \sum_{i=1}^{p}\phi_i^{(j)} x_{t-i}) + u_t.$$

Let

$$X = \begin{pmatrix} 1, x_{T-1},\ x_{T-2}, ..., x_{T-p} \\ 1, x_{T-2},\ x_{T-3}, ..., x_{T-p-1} \\ .... \\ 1,\ \ x_p, ; x_{p-1}, ..., x_1 \end{pmatrix}_{(T-p)\times(p+1)},$$

and

$$I_j(\gamma^0) = diag\left\{\Psi_T^{(j)}\left(\gamma^0\right), \Psi_{T-1}^{(j)}\left(\gamma^0\right), ..., \Psi_{p+1}^{(j)}\left(\gamma^0\right)\right\}.$$

Set $Y = (x_T, x_{T-1}, ..., x_{p+1})'$, $\beta_j = (\phi_0^{(j)}, \phi_1^{(j)}, ..., \phi_p^{(j)})'$, $j = 1, 2, 3, 4$, and $U = (u_T, u_{T-1}, ..., u_{p+1})'$.

Model (1) can be rewritten as the matrix form:

$$Y = \sum_{j=1}^{4} I_j(\gamma^0) X \beta_j + U. \tag{3}$$

Given $\gamma = (\gamma_1, \gamma_2)$, the OLS estimator for β_j is

$$\widehat{\beta}_j(\gamma) = (X' I_j(\gamma) X)^{-1} X' I_j(\gamma) Y, \ j = 1, 2, 3, 4, \tag{4}$$

where

$$I_j(\gamma) = diag\left\{\Psi_T^{(j)}(\gamma), \Psi_{T-1}^{(j)}(\gamma), ..., \Psi_{p+1}^{(j)}(\gamma)\right\}.$$

The residual sum of squares is

$$RSS_T(\gamma) = || \sum_{j=1}^{4} I_j(\gamma^0) X \beta_j + U - \sum_{j=1}^{4} I_j(\gamma) X \widehat{\beta}_j(\gamma) ||^2,$$

and

$$\widehat{\gamma} = \arg\min_{\gamma \in \Omega} RSS_T(\gamma), \tag{5}$$

where $\Omega = [\underline{\gamma_1}, \overline{\gamma_1}] \times [\underline{\gamma_2}, \overline{\gamma_2}]$.

The final structural estimators are then defined as

$$\widehat{\beta}_j(\widehat{\gamma}) = (X' I_j(\widehat{\gamma}) X)^{-1} X' I_j(\widehat{\gamma}) Y, \ j = 1, 2, 3, 4. \tag{6}$$

Theorem 1 *As $T \to \infty$ and under assumptions $A(1)$ to $A(4)$, we have $\widehat{\gamma} \xrightarrow{p} \gamma^0$ and $\widehat{\beta}_j(\widehat{\gamma}) \xrightarrow{p} \beta_j$, for $j = 1, 2, 3, 4$.*

Proof. See Appendix 1 ■

One can easily extend Theorem (1) to the restricted models in a similar vein.

4 Inference

4.1 Asymptotic Joint Distribution of $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$

Hansen (2000) derives the limit distribution of the threshold estimator in a single threshold model. Chong and Yan (2004) obtain the joint distribution of the threshold parameter vector for a two-regime threshold model with a similar method. In this section, we will derive the asymptotic approximation of the joint distribution of the least-squares estimator $\widehat{\gamma}$ for a four-regime M-TAR model under the assumption that the magnitude of change goes to zero at an appropriate rate.

To make the notation clear, we rewrite Model (3) as:

$$Y = X\beta_1 + \sum_{j=2}^{4} X_{\gamma^0}^{(j)} \delta^{(j)} + U, \tag{7}$$

where

$$X_{\gamma^0}^{(j)} = I_j\left(\gamma^0\right) X,\ j = 1, 2, 3, 4,$$

and

$$\delta^{(j)} = \beta_j - \beta_1,\ j = 2, 3, 4.$$

For any γ, we define

$$X_{\gamma}^{(j)} = I_j\left(\gamma\right) X,\ \ j = 1, 2, 3, 4.$$

Observe that: $X_{\gamma}^{(i)\prime} X_{\gamma}^{(j)} = 0$, if $i \neq j$, and $X' X_{\gamma}^{(j)} = X_{\gamma}^{(j)\prime} X_{\gamma}^{(j)}$, $j = 1, 2, 3, 4$.
Let $X_0^{(j)} = X_{\gamma^0}^{(j)}$, we have

$$X = \sum_{j=1}^{4} X_0^{(j)} = \sum_{j=1}^{4} X_{\gamma}^{(j)}.$$

The moment functionals are defined as:

$$M = E\left(X'X\right), \tag{8}$$

$$M_j\left(\gamma\right) = E\left(X_{\gamma}^{(j)\prime} X_{\gamma}^{(j)}\right),\ j = 1, 2, 3, 4, \tag{9}$$

$$D\left(\gamma\right) = E\left(X_t X_t' | z_t = \gamma\right), \tag{10}$$

$$V\left(\gamma\right) = E\left(X_t X_t' u_t^2 | z_t = \gamma\right), \tag{11}$$

where $X_t = (1, x_{t-1}, x_{t-2}, ..., x_{t-p})'$ and $t = p+1, p+2, ..., T$.

Set $D = D(\gamma_0)$, $V = V(\gamma_0)$, $D^* = (D, D, D)'$ and $V^* = (V, V, V)'$.

In addition to the assumptions $A(1)$ to $A(4)$, we make the following assumptions:

$(A5)$ Threshold variable $z_t = (z_{1t}, z_{2t})$ is i.i.d and follows the joint distribution $F(r)$. $F(r)$ is continuous and differentiable with respect to both variables. $f(r)$ denotes the joint density function of z and let $f = f(\gamma_0)$;

$(A6)$ For all $\gamma \in \Omega$ where $\Omega = [\underline{\gamma_1}, \overline{\gamma_1}] \times [\underline{\gamma_2}, \overline{\gamma_2}]$, $E\left(|X_t'U|^4 | z_t = \gamma\right) \leq a$ and $E\left(|X_t|^4 | z_t = \gamma\right) \leq b$ for some $a, b < \infty$, and $0 < f(\gamma) \leq \overline{f} < \infty$;

$(A7)$ $f(\gamma), D(\gamma)$ and $V(\gamma)$ are continuous at $\gamma = \gamma^0$;

$(A8)$ $\delta = (\delta^{(2)'}, \delta^{(3)'}, \delta^{(4)'})' = cT^{-\alpha}$, $0 < \alpha < \frac{1}{2}$, c is a constant vector;[8]

$(A9)$ $c' D^* c > 0$, $c' V^* c > 0$;

$(A10)$ $M_j(\gamma) > 0$ for all $\gamma \in \Omega$, where $\Omega = [\underline{\gamma_1}, \overline{\gamma_1}] \times [\underline{\gamma_2}, \overline{\gamma_2}]$, $j = 1, 2, 3, 4$.

When T is large enough, from assumption $(A4)$, we have:

(a) $\frac{1}{T}\sum_{t=1}^{T} I\left(z_{1t} \leq a, z_{2t} \leq b\right) \xrightarrow{p} F_1\left(a, b\right);$

(b) $\frac{1}{T}\sum_{t=1}^{T} I\left(z_{1t} \leq a, z_{2t} > b\right) \xrightarrow{p} F_2\left(a, b\right);$

(c) $\frac{1}{T}\sum_{t=1}^{T} I\left(z_{1t} > a, z_{2t} \leq b\right) \xrightarrow{p} F_3\left(a, b\right);$

(d) $\frac{1}{T}\sum_{t=1}^{T} I\left(z_{1t} > a, z_{2t} > b\right) \xrightarrow{p} F_4\left(a, b\right).$

In the neighborhood of the true threshold value, we have the following expansions:

$$M_i\left(\gamma\right) = M_i\left(\gamma^0\right) + \left(\gamma_1 - \gamma_1^0\right) DF_{i1}^0 + \left(\gamma_2 - \gamma_2^0\right) DF_{i2}^0 + o(1), \tag{12}$$

$$M_i\left(\gamma_1^0, \gamma_2\right) = M_i\left(\gamma^0\right) + \left(\gamma_2 - \gamma_2^0\right) DF_{i2}^0 + o(1), \tag{13}$$

[8] Because of the superconsistency of the threshold estimator, the distribution of the threshold estimator will degenerate to the true value for any fixed magnitude of change. To generate a meaningful distribution, the usual practice is to let the magnitude of change to go to zero at an appropriate rate.

$$M_i\left(\gamma_1,\gamma_2^0\right) = M_i\left(\gamma^0\right) + \left(\gamma_1-\gamma_1^0\right) DF_{i1}^0 + o(1), \tag{14}$$

where, $F_{i1}^0 = \frac{\partial F_i(\gamma)}{\partial \gamma_1}|_{\gamma=\gamma^0}$, $F_{i2}^0 = \frac{\partial F_i(\gamma)}{\partial \gamma_2}|_{\gamma=\gamma^0}$.

When $f(\gamma)$ is continuous at the $\gamma=\gamma^0$, we have,

$F_{11}^0 = F_{21}^0 = -F_{31}^0 = -F_{41}^0 = F_1^0 = \frac{\partial F(\gamma)}{\partial(\gamma_1)}|_{\gamma=\gamma^0}$;

$F_{12}^0 = F_{32}^0 = -F_{22}^0 = -F_{42}^0 = F_2^0 = \frac{\partial F(\gamma)}{\partial(\gamma_2)}|_{\gamma=\gamma^0}$.

Theorem 2 *Under the assumptions (A1) to (A10), if z_1 and z_2 are independent,*

$$\begin{aligned}
& T^{1-2\alpha}\frac{(c'D^*c)^2}{c'V^*c}\left((\widehat{\gamma}_1-\gamma_1^0)F_1^0,\ (\widehat{\gamma}_2-\gamma_2^0)F_2^0\right) \\
= \ & (\widehat{r}_1,\widehat{r}_2) \\
& \xrightarrow{d} \arg\min_{-\infty<r_1<\infty,-\infty<r_2<\infty}\left(\frac{|r_1|}{2}+W_1(r_1)+\frac{|r_2|}{2}+W_2(r_2)\right) \\
& \stackrel{d}{=} \arg\max_{-\infty<r_1<\infty,-\infty<r_2<\infty}\left(-\frac{1}{2}|r_1|+W_1(r_1)-\frac{1}{2}|r_2|+W_2(r_2)\right).
\end{aligned}$$

where

$$W_i(r) = \left\{\begin{array}{c} W(-r) \text{ if } r<0 \\ 0 \text{ if } r=0 \\ W(r) \text{ if } r>0 \end{array}\right\}, i=1,2.$$

$W(r)$ is a Brownian motion on $[0,\infty)$.

Proof. See Appendix 2 ■

Theorem 2 extends the work of Chong and Yan (2004). Their model is a restricted model with two regimes, i.e. $\delta_2=\delta_3=0$. We could obtain their result from Theorem 2 when $c=(0,0,\delta_4)'$.

Although the scale is different, $(\widehat{r}_1,\widehat{r}_2)$ follows the same distribution as that of

Chong and Yan (2004). Thus, the joint distribution has the same closed form.

$$\begin{aligned}
& F_{(\widehat{r}_1,\widehat{r}_2)}\left(a_1,a_2\right) \\
= \ & \Pi_{j=1}^2 F_{\widehat{r}_j}\left(a_j\right) \\
= \ & \Pi_{j=1}^2 \left(\begin{array}{c} 1+\sqrt{\frac{a_j}{2\pi}}\exp\left(-\frac{a_j}{8}\right)+\frac{3}{2}\exp\left(a_j\right)\Phi\left(-\frac{3\sqrt{a_j}}{2}\right) \\ -\frac{a_j+5}{2}\Phi\left(-\frac{\sqrt{a_j}}{2}\right) \end{array} \right),
\end{aligned} \tag{15}$$

and the joint density function can be found as

$$\begin{aligned}
& f_{(\widehat{r}_1,\widehat{r}_2)}\left(a_1,a_2\right) \\
= \ & \Pi_{j=1}^2 f_{\widehat{r}_j}\left(a_j\right) \\
= \ & \Pi_{j=1}^2 \frac{3}{2}\exp\left(a_j\right)\Phi\left(-\frac{3\sqrt{a_j}}{2}\right)-\frac{1}{2}\Phi\left(-\frac{\sqrt{a_j}}{2}\right),
\end{aligned} \tag{16}$$

where $\Phi\left(\cdot\right)$ is the cdf of a standard normal distribution, and $a_j > 0$.

In cases when some of the $a_j < 0$, we can replace those items in the above expression by $F_{\widehat{r}_j}\left(a_j\right) = 1 - F_{\widehat{r}_j}\left(-a_j\right)$ and $f_{\widehat{r}_j}\left(a_j\right) = f_{\widehat{r}_j}\left(-a_j\right)$.

Next, we consider the hypothesis

$$H_0 : \gamma = \gamma^0.$$

The Likelihood Ratio test of Hansen (1999, 2000) can be applied here.

Define

$$LR_T\left(\gamma_1^0, \gamma_2^0\right) = T\frac{RSS_T\left(\gamma_1^0,\gamma_2^0\right) - RSS_T\left(\widehat{\gamma}_1,\widehat{\gamma}_2\right)}{RSS_T\left(\widehat{\gamma}_1,\widehat{\gamma}_2\right)}. \tag{17}$$

H_0 is rejected for large $LR_T\left(\gamma_1^0,\gamma_2^0\right)$.

Under the assumption that threshold variables are independent, we have

$$LR_T\left(\gamma_1^0, \gamma_2^0\right) \xrightarrow{d} \eta^2\xi, \tag{18}$$

where

$$\xi = \xi_1 + \xi_2,$$

$$\xi_1 = \max_{-\infty<r_1<\infty} \left(-|r_1| + 2W_1(r_1)\right),$$

$$\xi_2 = \max_{-\infty<r_2<\infty} \left(-|r_2| + 2W_2(r_2)\right)$$

and

$$\eta^2 = \frac{c'V^*c}{\sigma^2 c'D^*c}.$$

Since u_t is a white noise sequence, we have $V^* = \sigma^2 D^*$, i.e., $\eta^2 = 1$.[9]
The distribution of ξ_i $(i = 1, 2)$ is

$$\Pr(\xi_i \leq x) = \left(1 - e^{-\frac{1}{2}x}\right)^2,$$

$$f_{\xi_i}(x) = \left(1 - e^{-\frac{1}{2}x}\right) e^{-\frac{1}{2}x}.$$

Thus,

$$\begin{aligned} \Pr(\xi \leq x) &= \Pr(\xi_1 + \xi_2 \leq x) \\ &= \int_0^x \Pr(\xi_1 \leq x - y) f_{\xi_2}(y)\, dy \\ &= 1 - (x+5)e^{-x} - 2(x-2)e^{-\frac{1}{2}x}. \end{aligned} \tag{19}$$

The density function is given by

$$f_\xi(x) = (x+4)e^{-x} + (x-4)e^{-\frac{1}{2}x} \tag{20}$$

Chong and Yan (2004) calculate the asymptotic critical value table of ξ from $m = 2$ to $m = 10$. When $m = 2$, the 95% critical value is 11.98.

[9] If we relax the assumption of white noise, the asymptotic distribution of $LR_T(\gamma_1^0, \gamma_2^0)$ will depend on the nuisance parameter η^2. To get the final critical value of $LR_T(\gamma_1^0, \gamma_2^0)$, we need to estimate η^2.

4.2 Testing Thresholds Effect: Model Selection Followed by Testing

Consider the null hypothesis that x_t is linear versus the alternative hypothesis that it is a threshold model. This problem has attracted much attention in recent years, partly due to the difficulty that thresholds γ are undefined under the null hypothesis. Hansen (1996) defines a likelihood ratio test as $J_T(\gamma) = \max_{\gamma\in\Gamma}(T(\widehat{\sigma}^2 - \widehat{\sigma}^2(\gamma))/\widehat{\sigma}^2(\gamma))$ and obtains its limiting distribution. The test considers the univariate case with two regimes and bootstrap method are used to obtain the critical value table. On the other hand, Petruccelli and Davis (1986) transform the problem into a standard test for break-point using the arranged regression approach. The test is quite general in that it does not require a prior specification of the threshold location. The cost, however, is that the power of the test is rather limited compared with that of the likelihood ratio tests. Besides, this method cannot be extended to the model with multiple threshold variables, where the arranged regression does not work.

In this section, we will generalize the likelihood ratio test statistics of Hansen (1996) to the M-TAR model.

As for model (7), the null hypothesis is:

$H_0 : \delta = 0,$

where $\delta = \begin{pmatrix} \delta^{(2)} \\ \delta^{(3)} \\ \delta^{(4)} \end{pmatrix}$.

The alternative is:

$H_1 : \delta \neq 0.$

The statistic is defined as:

$$J_T(\gamma) = \max_{\gamma\in\Omega}(T\frac{\widehat{\sigma}^2 - \widehat{\sigma}^2(\gamma)}{\widehat{\sigma}^2(\gamma)}), \tag{21}$$

where $\Omega = [\underline{\gamma_1}, \overline{\gamma_1}] \times [\underline{\gamma_2}, \overline{\gamma_2}]$.

$\widehat{\sigma}^2$ is the residual sum of squares under the null hypothesis, while $\widehat{\sigma}^2(\gamma)$ is the residual sum of squares under the alternatives. The limiting distribution of $J_T(\gamma)$ typically depends on the unknown γ and model specific moments and cannot be tabulated, thus the bootstrap method is used to calculate the critical values.

If the null cannot be rejected, we conclude that there is no threshold effect for both threshold variables and the number of regimes is one. If the null hypothesis is rejected, we conclude that there is a threshold effect for at least one of the variables. However, we still need to estimate the number of regimes. The OLS estimation for the threshold variables is: $\widehat{\gamma} = \arg\min_{\gamma\in\Omega} \widehat{\sigma}^2(\gamma) = \arg\max_{\gamma\in\Omega} J_T(\gamma)$. Under the alternative hypothesis, $\widehat{\gamma}$ is consistent.[10]

The following steps are for deciding the number of regimes.

The first step is to test a three-regime model against a four-regime model. The hypothesis can be separated into the following sub-hypotheses:

(I) $H_0 : \delta^{(2)} = 0$;

(II) $H_0 : \delta^{(3)} = 0$;

(III) $H_0 : \delta^{(4)} = 0$;

(IV) $H_0 : \delta^{(2)} = \delta^{(3)}$;

(V) $H_0 : \delta^{(2)} = \delta^{(4)}$;

(VI) $H_0 : \delta^{(3)} = \delta^{(4)}$.

The alternative is:

H_1 : the model is a four-regime model.

Six related likelihood ratio tests could be defined specifically to test these hypotheses:

$$J_T(\widehat{\gamma}) = T\frac{[\widehat{\sigma}_0^2(\widehat{\gamma}) - \widehat{\sigma}_1^2(\widehat{\gamma})]}{\widehat{\sigma}_1^2(\widehat{\gamma})}, \tag{22}$$

where $\widehat{\sigma}_0^2(\widehat{\gamma})$ is the residual sum of squares under the null hypothesis, while $\widehat{\sigma}_1^2(\widehat{\gamma})$ is the residual sum of squares under the alternatives.

If the null is rejected for each sub-hypothesis, then we conclude that there are four regimes. As long as one of the hypotheses is accepted, there are less than four regimes and we go to the next step.

The second step is to test a two-regime model against a three regime model. Based on the result of the first step, we still need to design three sub-hypotheses.

[10]Refer to Theorem 3.

For illustration, assuming (I) $H_0 : \delta^{(2)} = 0$ is accepted in the first step, we should go on testing whether there is another shift between other regimes.

The hypotheses are:

(II) $H_0 : \delta^{(3)} = \delta^{(2)} = 0$;

(III) $H_0 : \delta^{(4)} = \delta^{(2)} = 0$;

(VI) $H_0 : \delta^{(3)} = \delta^{(4)}, \delta^{(2)} = 0$.

The alternative is :

$H_1 : \delta^{(2)} = 0$.

$J_T(\widehat{\gamma})$ could be used to test the above three hypothese. If all of the null hypotheses are rejected, we conclude that the model is a three-regime model. As long as one null is accepted, we have to go to the next step to test the hypothesis of a one-regime model against a two-regime model. The number of regimes can be determined by this sequential procedure.

Since the asymptotic distributions of the above likelihood ratio tests are non-standard, we adopt the bootstrap method of Hansen (1999) to calculate the critical values. Firstly, the residuals under H_1 will be used as the empirical distribution from which a sample of size T with replacement will be drawn and used to create a bootstraps sample under H_0. The regressors and threshold variables will be fixed in the repeated bootstrap samples. The structural parameters and threshold value will also be fixed at their estimated values under H_0. Repeating this procedure a large number of times and calculating the percentage of draws for which the simulated statistic exceeds the one obtained from the original sample, we get the bootstrapping p-value under H_0. The null will be rejected when the p-value is too small. Hansen (1999) has shown that the test is consistent.

The consistency of the test in each step implies that the probability to reject the null under the alternative converges to 1. Thus, the under-estimation of the number of regimes can be solved in a large sample. The problem is the over-estimation of the number of regimes, since there is always a chance of rejecting the null of no threshold effect even it is true. The number of regimes depends on the probability of detecting the shift changes successfully. In other words, it depends on the size of the test α. In general, if there are s ($s \leq 4$ for two threshold variables) regimes, before the step to test s against $s-1$ regimes, we should accept $(4-s)$ times of null and the probability for over-estimation is:

$$P(\widehat{s} > s) \overset{T\to\infty}{\to} \alpha \sum_{j=1}^{4-s} (1-a)^{j-1} = 1 - (1-\alpha)^{4-s} \tag{23}$$

and

$$P(\widehat{s} = s) \overset{T\to\infty}{\to} 1 - \alpha \sum_{j=1}^{4-s} (1-a)^{j-1} = (1-\alpha)^{4-s}. \tag{24}$$

The consistency of the $\widehat{s}$ can be obtained if we let α to go to zero at a suitable rate, for example, let $\alpha = \frac{1}{\log T}$. Therefore, we can find the number of regimes consistently through the sequential test.

5 Modeling

Identifying an adequate M-TAR model for a given data set involves selecting parameters such as threshold values, structural parameters and the order of AR model in each regime. In some early applications, Tong and Lim (1980) and Tsay (1989) utilize the past experience and substantive information, such as the value of t-test or graph to determine the threshold location and the number of regimes. The method depends on subjective judgment. Recently, Tsay (1998) uses a grid search method to locate the threshold and select the order in each regime simultaneously by minimizing the sum of AIC for all regimes. He assumes the number of regime to be known. Moreover, the method could not obtain consistent threshold estimators and the computational workload is very huge when there are more than one threshold variable.

In this section, we present a sequential modeling procedure for M-TAR model. The procedure not only can obtain consistent threshold estimators but also save the workload of computation compared with Tsay (1998)' s method. We first show the consistency of the least squares estimators in a misspecified threshold model.

5.1 Generic Consistency of the Threshold Estimators under Specification Errors

As is well known, break-point estimators are still consistent in a structural change model with misspecification in the functional form of the regression equation (Chong, 2003). Furthermore, the consistency can be guaranteed even when misspecifications in the functional form and the number of break changes are both present (Bai and Chong et al., 2004). The structural-change model can be considered as a special case of the threshold model by using time as the threshold variable. Reversely, the threshold model with a single threshold variable can be converted into a standard break-point model by re-indexing the threshold variable (Tsay, 1998). Thus, the result of the structural-change model can be extended to the threshold model. However, once the threshold model is allowed to have more than one threshold variable, the above transformation does not work. The related results of the threshold model with multiple threshold variables have yet to be developed.

Consider a threshold model with two threshold variables and two regimes:[11]

$$Y = I_1(\gamma^0)F\beta_1 + [I - I_1(\gamma^0)]F\beta_2 + U. \tag{25}$$

The estimated model misspecifies the number of regimes and the functional form:[12]

$$Y = \sum_{i=1}^{4} I_i(\gamma)G\widehat{\beta}_i + \widehat{U}.$$

Y is a T by 1 matrix with elements y_t, $t = 1, 2, ..., T$.

F is a T by P matrix with the $(t,p)th$ element $f_p(x_{tp})$, where $f_p(.)$ is a real value function. Note that F could be the lag values of Y, thus the model includes time series models.

G is a T by L matrix with the $(t,l)th$ element $f_l(x_{tl})$, where $f_l(.)$ is a real value function.

---

[11] The model can be extended to a model with three regimes.

[12] Note that the estimated model is a full model so that the assumed number of regimes is larger or equal than the actual number of regimes, but this condition is not a necessary condition for the consistency of the threshold estimators.

Let Q_{ff} and Q_{gg} be positive definite and non-stochastic matrices, which are defined as the covariance matrices of F and G respectively. Let Q_{fg} be a non-stochastic matrix defined as the covariance matrix between F and G. We need to assume that $Q_{gf}(\beta_1 - \beta_2) \neq 0$, if $\beta_1 \neq \beta_2$.[13]

U is a T by 1 matrix with elements u_t, where u_t is i.i.d. with zero means and variance σ^2, and finite forth order moment.

Given the threshold values γ, the OLS estimators for β can be written as

$$\widehat{\beta}_i(\gamma) = \left(G' I_i(\gamma) G\right)^{-1} G' I_i(\gamma) Y.$$

The residual sum of squares is:

$$RSS_T(\gamma) = \left(Y - \sum_{i=1}^{4} I_i(\gamma) G \widehat{\beta}_i(\gamma)\right)' \left(Y - \sum_{i=1}^{4} I_i(\gamma) G \widehat{\beta}_i(\gamma)\right),$$

and the threshold estimators can be defined as

$$\widehat{\gamma} = \arg\min_{\gamma \in \Omega} RSS_T(\gamma),$$

where $\Omega = [\underline{\gamma_1}, \overline{\gamma_1}] \times [\underline{\gamma_2}, \overline{\gamma_2}]$.

Theorem 3 *In a misspecified threshold model, if $Q_{gf}(\beta_1 - \beta_2) \neq 0$, then under H_1 : $\beta_1 \neq \beta_2$, $\widehat{\gamma} \xrightarrow{p} \gamma^0$.*

Proof. See Appendix 3. ■

Theorem 3 can be extended to an M-TAR model with misspecifications in the number of regimes and the order in each regime.

[13] The condition is necessary for the result. Otherwise, we will fail to detect the threshold effect. We refer to this problem as the "wiping effect" caused by the misspecification. The probability of the failure depends on the degree of the misspecification. For example, the probability equals one when $cov(G, F) = 0$. Generally, if G contains less information about F, the probability will be larger. In turn, if G includes all of the variables undergoing threshold change, and there exists a nonzero correlation between at least one pair of functional forms, then this failure will not happen. However, since both G and F are functions of X, the chance that $(\beta_1 - \beta_2)$ is the solution of equation $Q_{gf}X = 0$ is very small. Therefore, the assumption is easily satisfied.

Consider an M-TAR model with two regimes:

$$Y = I_1(\gamma^0)X_{(T-p)\times(p+1)}\beta_1 + [I - I_1(\gamma^0)]X_{(T-p)\times(p+1)}\beta_2 + U, \tag{26}$$

where I is an identity matrix, and

$$I_1(\gamma^0) = diag\left\{\Psi_T\left(\gamma^0\right), \Psi_{T-2}\left(\gamma^0\right), ..., \Psi_{p+1}\left(\gamma^0\right)\right\},$$

$$Y = (x_T, x_{T-1}, ..., x_{p+1})'_{(T-p)\times 1}, X_{(T-p)\times(p+1)} = \begin{pmatrix} 1, & x_{T-1}, & x_{T-2}, ...x_{T-p} \\ 1, & x_{T-2}, & x_{T-3}, ...x_{T-p-1} \\ & .... & \\ 1, & x_p, & x_{p-1}, ...x_1 \end{pmatrix}_{(T-p)\times(p+1)}$$

and $\Psi_t(\gamma) = I\left(z_{1t} \leq \gamma_1, z_{2t} \leq \gamma_2\right)$ is an indicator function.

The estimated model is misspecified as

$$Y = [I_1(\gamma), I_2(\gamma), I_3(\gamma), I_4(\gamma)]X_{(T-q)\times(q+1)} \cdot \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \widehat{\beta}_3 \\ \widehat{\beta}_4 \end{pmatrix} + \widehat{U}, \tag{27}$$

$$I_j(\gamma) = diag\left\{\Psi_T^{(j)}\left(\gamma\right), \Psi_{T-1}^{(j)}\left(\gamma\right), ..., \Psi_{q+1}^{(j)}\left(\gamma\right)\right\},$$

where $X_{(T-q)\times(q+1)}$ is defined as same as $X_{(T-p)\times(p+1)}$ and $p \neq q$. The estimated model has misspecifications in the number of regimes and the order in each regime.

Q_{pp} and Q_{qq} are positive definite and non-stochastic matrices, which are defined as the covariance matrix of $X_{(T-p)\times(p+1)}$ and $X_{(T-q)\times(q+1)}$ respectively. Q_{pq} is a non-stochastic matrix defined as the covariance matrix between $X_{(T-p)\times(p+1)}$ and $X_{(T-q)\times(q+1)}$. We assume that $Q_{pq}(\beta_1 - \beta_2) \neq 0$ if $\beta_1 \neq \beta_2$.

Given the threshold value γ, the OLS estimators for β can be written as

$$\widehat{\beta}_i(\gamma) = \left(X'_{(T-q)\times(q+1)}I_i(\gamma)X_{(T-q)\times(q+1)}\right)^{-1} X'_{(T-q)\times(q+1)}I_i(\gamma)Y.$$

The residual sum of squares is

$$RSS_T(\gamma) = \left(Y - \sum_{i=1}^{4} I_i(\gamma) X_{(T-q)\times(q+1)} \widehat{\beta}_i(\gamma)\right)' \left(Y - \sum_{i=1}^{4} I_i(\gamma) X_{(T-q)\times(q+1)} \widehat{\beta}_i(\gamma)\right),$$

and the threshold estimators are defined as

$$\widehat{\gamma} = \arg\min_{\gamma \in \Omega} RSS_T(\gamma),$$

where $\Omega = [\underline{\gamma_1}, \overline{\gamma_1}] \times [\underline{\gamma_2}, \overline{\gamma_2}]$.

Corollary 4 *In a misspecified threshold AR model, if $Q_{pq}(\beta_1 - \beta_2) \neq 0$, then under $H_1 : \beta_1 \neq \beta_2$, $\widehat{\gamma} \xrightarrow{p} \gamma^0$.*

Proof. In the proof of Theorem 3, we just need to let $f_i(x_t) = g_j(x_t) = x_t$ for any i and j. F is a $(T-p) \times (p+1)$ matrix and G is a $(T-q) \times (q+1)$ matrix. This completes the proof of the Corollary. ■

5.2 Modeling Procedure

We now discuss the modeling procedure.

Step 1: We first estimate an M-TAR model

$$x_t = \sum_{i=1}^{4} (\widehat{\beta}_{i0} + \widehat{\beta}_{i1} x_{t-1}) \Psi_t^{(i)}(\gamma) + \widehat{e}_t,$$

and get $\widehat{\gamma}_T = \arg\min_{\gamma \in \Omega} RSS_T(\gamma)$, where $\Omega = [\underline{\gamma_1}, \overline{\gamma_1}] \times [\underline{\gamma_2}, \overline{\gamma_2}]$.

Step 2: perform the sequential likelihood ratio test described in Section 4.2 to determine the number of regimes.

Step 3: based on the threshold values and the number of regimes derived from Step 1 and Step 2, use AIC rules to select the order in each regime.[14]

In our case,

$$AIC_i(p_i) = n_i \ln[RSS_i(\widehat{\gamma}_T)/n_i] + 2(p_i + 1), \tag{28}$$

[14]The AIC has been used in the literature to select threshold autoregressive models (see Tsay 1998). Given some conditions, AIC is asymptotically equivalent to selecting the model that has the smallest generalized residual variance using the conditional least squares method.

n_i : the number of the observations in the ith regime;

p_i : the order of the model in the ith regime;

$RSS_i(\widehat{\gamma}_T)$: the residual sum of squares in the ith regime.

Define the order estimator as

$$\widehat{p}_i = \arg\min_{1 \leq p_i \leq 10} [AIC_i(p_i)]. \tag{29}$$

Here we use 10 as the maximum order considered in the model.[15]

The AIC for the whole model can be written as

$$NAIC = \sum_{i=1}^{s} AIC_i(\widehat{p}_i)/T, \tag{30}$$

T is the number of effective observations and s is the number of regimes.

Step 4: use the model obtained from Step 3 to refine the threshold values.

6 Monte Carlo Simulations

In this section, Monte Carlo simulations will be performed to demonstrate the above results. All simulations are done by R-Language programs, which are available from the authors upon request.

Experiment A:

In this experiment, we will show the consistent estimation for threshold models with various kinds of misspecifications. The cases with dependent threshold variables are also included.

Sample size: $T = 200$;

Number of replications: $N = 500$;

$e_t \sim N(0,1)$, $\varepsilon_t \sim N(0,1)$, $z_{1t} \sim N(0,1)$.

We simulate three data generating processes:

DGP 1: $y_t = x_t^2 I(z_{1t} < 0 \text{ or } z_{2t} < 0) - 2x_t^2 I(z_{1t} \geq 0 \text{ and } z_{2t} \geq 0) + e_t$, where $x_t \sim U(0, 10)$;

[15]This value depends on the requirement in practice.

DGP 2 : $x_t = (0.3x_{t-1}+0.3x_{t-2})I(z_{1t} < 0 \text{ or } z_{2t} < 0)+(-0.3x_{t-1}-0.3x_{t-2})I(z_{1t} \geq 0$ and $z_{2t} \geq 0) + e_t$;

DGP 3 : $x_t = 0.3x_{t-1}I(z_{1t} < 0 \text{ or } z_{2t} < 0) - 0.3x_{t-1}I(z_{1t} \geq 0 \text{ and } z_{2t} \geq 0) + e_t$,

Three misspecified models are estimated respectively:

Model A: $y_t = \sum_{i=1}^{4} \widehat{\beta}_i x_t \Psi_t^{(i)}(\gamma) + \widehat{e}_t$.

Model B: $x_t = \sum_{i=1}^{4} \widehat{\beta}_i x_{t-1} \Psi_t^{(i)}(\gamma) + \widehat{e}_t$.

Model C: $x_t = \sum_{i=1}^{4} (\widehat{\beta}_{1i} x_{t-1} + \widehat{\beta}_{2i} x_{t-2}) \Psi_t^{(i)}(\gamma) + \widehat{e}_t$.

The estimation results are shown in Table 1:

Table 1 The Estimated Result for Experiment A

DGP	Estimated Model	z_{2t}	$\overline{\widehat{\gamma}_1}$	$Var(\widehat{\gamma}_1)$	$\overline{\widehat{\gamma}_2}$	$Var(\widehat{\gamma}_2)$
1	A	$N(0,1)$	0.001	0.0004	0.001	0.0003
1	A	$z_{1t} + \varepsilon_t$	−0.002	0.0003	0.002	0.0001
2	B	$N(0,1)$	0.005	0.0300	−0.001	0.026
2	B	$z_{1t} + \varepsilon_t$	−0.004	0.0310	−0.003	0.029
3	C	$N(0,1)$	−0.001	0.0250	−0.006	0.034
3	C	$z_{1t} + \varepsilon_t$	−0.006	0.0310	0.002	0.032

Table 1 shows that the estimators are consistent under the existence of various misspecifications.

Experiment B:

This experiment shows the performance of the modeling procedures.

DGP is a three-regime threshold autoregressive model:

$x_t = \beta_1 x_{t-1} + \beta_2 x_{t-2} - (\beta_1^{(1)} x_{t-1} + \beta_2^{(1)} x_{t-2}) I(z_{1t} \leq \gamma_1^0 \text{ and } z_{2t} \leq \gamma_2^0) + (\beta_1^{(2)} x_{t-1} + \beta_2^{(2)} x_{t-2}) I(z_{1t} > \gamma_1^0 \text{ and } z_{2t} > \gamma_2^0) + e_t$,

where $z_{1t} \sim N(0,1)$, $\varepsilon_t \sim N(0,1)$, $z_{2t} = z_{1t} + \varepsilon_t$;

$\beta_1 = \beta_2 = 0.4$, $\beta_0^{(1)} = \beta_1^{(1)} = 0.8$, $\beta_0^{(2)} = \beta_1^{(2)} = 0.4$, and $\gamma_1^0 = \gamma_2^0 = 0$.

Sample size $T = 1000$.

Step 1: locate the threshold variables by the OLS method. We estimate a full threshold autoregressive model with AR(1) in each regime. Let $q_{(j)}$ denote the $(j/N)th$ quantile of the range $(\underline{\gamma}, \overline{\gamma})$. Here, we set $N = 100$, and $(\underline{\gamma}, \overline{\gamma}) = (-0.5, 0.5)$.

The estimated results are $\widehat{\gamma}_1^0 = -0.03$ and $\widehat{\gamma}_2^0 = 0.01$, which are very close to the true values.

Step 2: use the likelihood ratio tests described in Section 4.2 to choose the appropriate number of regimes.

Table 2(a) shows the results for a three-regime model against a four-regime model;

Table 2(b) shows the result for a two-regime model against a three-regime model;

Table 2(a) The result for a three-regime model against a four-regime model

	$H_1 : s = 4$	
	$H_0 : s = 3$ $(\delta_2 = 0)$	$H_0 : s = 3$ $(\delta_2 = \delta_3)$
$J_T(\gamma)$	81.39	1.13
p-value	0.00	0.975

Table 2(b) The result for a two-regime model against a three-regime model

	$H_1 : s = 3(\delta_2 = \delta_3)$		
	$H_0 : s = 2$ $(\delta_2 = 0)$	$H_0 : s = 2$ $(\delta_4 = \delta_3)$	$H_0 : s = 2$ $(\delta_4 = 0)$
$J_T(\widehat{\gamma})$	112.05	77.07	45.6
p-value	0.00	0.00	0.00

The results indicate that the model with three regimes best fits the data.

Step 3: based on the results from Step 1 and Step 2, AIC rules are applied here to obtain the order in each regime. The results are indicated in Table 3.

Table 3 The AIC in each regime

Order	$AIC(p_1)$	$AIC(p_2)$	$AIC(p_3)$
1	77.27	28.93	22.36
2	40.09	18.93	−37.44
3	41.79	20.03	−39.37
4	43.78	22.01	−41.71
5	45.19	23.45	−40.61
6	47.18	24.13	−38.82
7	47.88	26.12	−36.96
8	50.39	28.11	−35.06
9	52.58	28.08	−32.98
10	55.25	26.5	−31.97

Take $\widehat{p}_i$ as

$$\widehat{p}_i = \arg \min_{1 \leq p_i \leq 10} [AIC_i(p_i)]. \tag{31}$$

The estimated results are:

$$x_t = \left\{ \begin{array}{c} -0.04 - 0.37x_{t-1} - 0.39x_{t-2} \text{ when } z_{1t} < \widehat{\gamma}_1^0 \text{ or } z_{2t} < \widehat{\gamma}_2^0 \\ 0.09 + 0.43x_{t-1} + 0.43x_{t-2} - 0.09x_{t-3} + 0.09x_{t-4} \text{ when } z_{1t} \geq \widehat{\gamma}_1^0 \text{ and } z_{2t} \geq \widehat{\gamma}_2^0 \\ -0.07 + 0.09x_{t-1} + 0.14x_{t-2}, \text{ otherwise} \end{array} \right\}$$

Step 4: use the model obtained from Step 3 to refine the thresholds' value.

The final result are $\widehat{\gamma}_1^0 = -0.01$ and $\widehat{\gamma}_2^0 = 0$.

7 Empirical Application in the Financial Market

It is a belief in the market that the behavior of stock price is consistent with a nonlinear data-generating process. A growing body of studies on single threshold variable model has been proposed to describe and explain the nonlinearity found in the stock price series. For example, Tong (1983) uses the STAR model to fit the IBM stock price. Li and Lam (1995) investigate the asymmetric behavior of stock prices in bear and bull markets by using a threshold type ARCH model. Recently Shively

(2003) employs a three-regime threshold random-walk model to fit the daily data of CAC(40)[16], DAX(30)[17] and other indices.

In this section, we apply the M-TAR model to Hang Seng Index (HSI), with the past values of price and turnover as threshold variables. To the best of our knowledge, it is the first study using the turnover as a threshold variable. Turnover is used as it provides insight into the structure of market. It is an adage on Wall Street that turnover is relatively heavy in a bull market and light in a bear market. Such asymmetry in the volume relationship is well-known (for example, Karpoff, 1987). Secondly, price changes are usually interpreted as the market evaluation of new information, while the corresponding turnover is considered as an indicator of the extent to which investors agree with the information. For example, a huge rise in price with a small turnover is usually not a reliable indicator to predict the future boom of the market.

The two threshold variables are defined as.

$$Rap_t = \frac{PMA20_t}{PMA250_t}, \tag{32}$$

$$Rav_t = \frac{VMA20_t}{VMA250_t}, \tag{33}$$

where $PMA250_t$ and $VMA250_t$ are average price and turnover for past 250 trading days, while $PMA20_t$ and $VMA20_t$ are average price and turnover for past 20 days. The 250-day average line of price is usually considered as the boundary between a bull and a bear market by many investors. The 20-day average lines are used here to smooth variables. When $PMA20_t$ is above $PMA250_t$, i.e., $Rap_t > 1$, then we define the market as a bull market. Otherwise, it is a bear market.

The threshold model can be written as

$$x_t = \sum_{j=1}^{4} \Psi_t^{(j)} \left(\gamma^0\right) (\phi_0^{(j)} + \phi_1^{(j)} x_{t-1} + \phi_2^{(j)} x_{t-2}, ..., + \phi_p^{(j)} x_{t-p}) + \varepsilon_t, \tag{34}$$

[16]The stock index of France.

[17]The stock index of Germany.

where

$$\Psi_t^{(1)}(\gamma^0) = I\left(Rap_t \le \gamma_1^0,\ Rav_t \le \gamma_2^0\right),$$
$$\Psi_t^{(2)}(\gamma^0) = I\left(Rap_t \le \gamma_1^0,\ Rav_t > \gamma_2^0\right),$$
$$\Psi_t^{(3)}(\gamma^0) = I\left(Rap_t > \gamma_1^0,\ Rav_t \le \gamma_2^0\right),$$
$$\Psi_t^{(4)}(\gamma^0) = I\left(Rap_t > \gamma_1^0,\ Rav_t > \gamma_2^0\right).$$

7.1 Data Description

The data set is the daily return series of the non-dividend Hang Seng Index (HSI) from Jun. 3rd 1995 to Jan. 13th 2005.[18] HSI is a weighted average index of blue chips' prices in the Hong Kong stock market. The return series is defined as the log-difference of the index. The turnover series is the daily turnover of the whole market.[19] There are about 2500 observations in the sample. Since the first 250 data are used to calculate the moving average line, the number of the effective observations is about 2250.

Figure 1 and Figure 2 show the data and their moving average lines.

FIGURE 1 HERE

FIGURE 2 HERE

Figure 3 shows the two threshold variables Rap_t and Rav_t.

FIGURE 3 HERE

7.2 Estimated Results

In this section, we will use the modeling procedure described in Section 5.2 to estimate the model.

[18]The daily turnover of the market is only available from Jun. 1995 in CEIC database.

[19]In principle, we should use the turnover of blue chips here, but it is not available in the CEIC database. We thus take the turnover of the whole market as an instrument variable of the turnover of blue chips.

The estimated threshold values of Step 1 are: $\widehat{\gamma}_{rap} = 0.96$ and $\widehat{\gamma}_{rav} = 1.14$, which are very close to one.

We apply the sequential likelihood ratio test described in Section 4.2 to decide the number of regimes. The test results are shown in Table 4.

Table 4 The likelihood ratio test result

	H_1: $s=4$					
	H_0: $s=3$ ($\delta_2=0$)	H_0: $s=3$ ($\delta_2=\delta_3$)	H_0: $s=3$ ($\delta_4=0$)	H_0: $s=3$ ($\delta_3=0$)	H_0: $s=3$ ($\delta_2=\delta_4$)	H_0: $s=3$ ($\delta_3=\delta_4$)
$J_T(\gamma)$	102.5	67.83	16.03	16.78	99.68	24.93
p-value	0.00	0.00	0.05	0.05	0.00	0.025

The results indicate the four-regime model best fits the return series.

Table 5 shows the final estimated result and the threshold estimations are refined to be: $\widehat{\gamma} = (0.98, 1.14)$

Table 5 The estimated threshold model for the return series of Hang Seng Index

Model	The model parameters
Four-regime	$p_t = -0.0002 + 0.067p_{t-1} - 0.074p_{t-2} + 0.047p_{t-3} - 0.008p_{t-4} - 0.082p_{t-5}$, if $Rap_t < 0.98$ and $Rav_t < 1.14$;
	$p_t = 0.002 - 0.47p_{t-1} - 0.091p_{t-2} + 0.077p_{t-3} - 0.302p_{t-4} - 0.365p_{t-5} - 0.259p_{t-6} - 0.125p_{t-7}$, if $Rap_t < 0.98$ and $Rav_t > 1.14$;
	$p_t = -0.0001 + 0.019p_{t-1}$, if $Rap_t > 0.98$ and $Rav_t < 1.14$
	$p_t = 0.0006 + 0.196p_{t-1} - 0.178p_{t-2} + 0.088p_{t-3}$, if $Rap_t > 0.98$ and $Rav_t > 1.14$

To see the adequacy of the model, we consider the hypothesis that whether the residual is white noise or not. Figure 4 shows the residual for the model and Figure 5 indicates the PACF and ACF respectively.

FIGURE 4 HERE

FIGURE 5 HERE

Table 6 shows the exact values of ACF and PACF and Ljung-Box Q-statistics:

Table 6 ACF and PACF of residuals

Autocorrelation Functions(ACF)

Lag 1 :	0.0024	0.0199	0.0251	0.0002	0.0076	0.0306
7 :	0.0113	−0.0005	0.0190	0.0300	0.0102	0.0108
13 :	0.0364	0.0019	0.0273	−0.008	0.0023	−0.006
19 :	−0.011	0.0402	−0.011	0.0050	0.0027	−0.024

Partial Autocorrelation Functions (PACF)

Lag 1 :	0.0024	0.0199	0.0250	−0.0002	0.0066	0.0300
7 :	0.0109	−0.002	0.0172	0.0295	0.0091	0.0077
13 :	0.0341	0.0009	0.0242	−0.012	0.0002	−0.009
19 :	−0.014	0.0383	−0.012	0.0024	−0.0005	−0.025

Ljung-Box Q-Statistics

$Q(8) = 4.610$. Significance Level 0.7982

$Q(16) = 12.386$. Significance Level 0.7169

$Q(24) = 17.892$. Significance Level 0.8081

In both tables and figures, the four-regime model appears adequate since there is no significant Q value being found. Both PACF and ACF fail to reject the null hypotheses at the 5-percent significance level. The results show that the residuals are white noises.

Our estimated results indicate the return series should be separated into four regimes. This conclusion is consistent with several previous studies of stock market. Granville (1965) divides the market index process into four regimes depending on the On Balance Volume (OBV) and the past price; Lee and Swaminathan (2000) classify stocks into four categories according to the trading volume and the price performance.

Based on this result, a four-regime cycle could be designed to explain the nonlinearity of the return series:

Regime I: the first stage of the bull market. Both *Rap* and *Rav* are above their thresholds. In this regime, price usually rises very quickly and trading volume is very large.

Regime II: the second stage of the bull market. *Rap* is still above its threshold but *Rav* drops to the level below its threshold. A small turnover implies that the ascending momentum is slowing down and the end of the bull market is coming.

Regime III: the first stage of the bear market. *Rap* is below its threshold but *Rav* rises to the level above the threshold. A large turnover implies a large descending momentum as more investors sell their stocks. The market drops dramatically in this regime.

Regime IV: the second stage of the bear market. *Rap* is still below its threshold and *Rav* falls to the level under the threshold. A small turnover indicates that selling pressure decreases compared with the first stage of the bear market. The market starts to recover.

Figure 6 illustrates some of the more salient features of our empirical findings.

<u>**FIGURE 6 HERE**</u>

Table 7 indicates the regimes distribution of Hang Seng Index from Jun.1995 to Jan.2005.

Table 7 The Regimes Distribution for Hang Seng Index from Jun. 1996 to Jan. 2005

Period	Regime	Period	Regime
(10/06/96, 24/09/96)	R II	(12/10/01, 11/06/01)	R IV
(25/09/96, 9/10/97)	R I	(12/06/01, 26/06/01)	R III
(13/10/97, 24/10/97)	R II	(27/06/01, 5/05/02)	R IV
(28/10/97, 19/11/97)	R III	(8/05/02, 5/06/02)	R I
(20/10/97, 3/11/98)	R IV	(6/06/02, 28/06/02)	R II
(4/11/98, 27/11/98)	R I	(2/07/02, 28/04/03)	R IV
(30/11/98, 16/04/99)	R II	(29/04/03, 11/06/03)	R III
(19/04/99, 2/06/99)	R I	(12/06/03, 8/04/04)	R I
(3/06/99, 22/06/99)	R II	(9/04/04, 4/01/05)	R II
(23/06/99, 7/04/00)	R I	(5/01/05, 13/01/05)	R I
(10/04/00, 11/10/00)	R II		

From Table 7, we could find a complete cycle from Sep. 1996 to Nov. 1998. Meanwhile, there are two cycles during Jun. 1999 and Jun. 2003. The average length for a cycle is two years.

8 Conclusion

Threshold autoregressive models capture the nonlinear properties of a time series and can be widely applied to model economic time series. Traditional TAR models only allow for a single threshold variable. This thesis examines a new kind of threshold autoregressive model that allows multiple threshold variables. Different from the work of Chong and Yan (2004), we set the number of threshold variables to two but leave the number of regimes unknown. Two threshold variables can split the sample into four categories at most. We refer to the threshold model as the full model when shifts occur among all of the categories. Certain restrictions on structural parameters are allowed so that changes just occur in a subset of regimes, and then the model becomes a restricted model, with just two regimes or three regimes.

The theoretical contributions of this thesis are twofold. First, we investigate the statistical properties of the OLS estimators and derive their asymptotic distribution. In a full model, it is shown that, under some regular conditions, the least squares threshold estimators are strongly consistent. With a similar approach used by Hansen (2000) and Chong and Yan (2004), we have derived the asymptotic approximation to the joint distribution of the least-squares estimator $\widehat{\gamma}$ of the threshold parameter vector γ. The distribution has a similar form as that of Chong and Yan (2004) but with a different scale. In particular, for the case where thresholds are independent, the joint asymptotic distribution of the thresholds estimator $\widehat{\gamma}$ is a combination of the distribution functions of a single threshold model. A likelihood ratio test is also designed to detect the threshold effect and choose the specific model. Second, we suggest an applicable modeling procedure based on the consistent estimators for a misspecified threshold model. The sequential model-building procedure consists of four steps, with AIC being used to decide the order of the model. The procedure not only saves the workload of computation in empirical studies, but also obtains consistent threshold estimators. Monte Carlo simulations are presented to highlight the performance of the modeling procedure and the test in the finite-sample case.

We apply our model to the return series of Hang Seng Index. The model uses past values of price and turnover as threshold variables. To the best of our knowledge, this work is the first one to examine the return series by using turnover as a threshold variable. Many threshold models have been applied to study the behavior of the return series, but they only take the lag value of the return as the single threshold variable and neglect the role of the turnover. Empirical studies have shown that turnover has a close relation with the return series and is helpful to classify the market structure. Therefore, our model reflects more information than previous studies by incorporating turnover into the model as a threshold variable. The model detects a threshold effect in the return series and suggests a four-regime cycle which provides a possible explanation to the nonlinearity of the series.

An interesting extension of our model is to allow each threshold variable to have more than one threshold. This model can deal with more general cases in reality and has a wider application. However, it is beyond the scope of this thesis and left for further research.

REFERENCES

1. Bai, J., T.L. Chong, X. Wang and H.Q. Chen (2004), Generic Consistency of the Break-Point Estimators under Specification Errors in a Multiple-Break Model. *Working paper.*

2. Chan, K. S. and H. Tong (1986), On Estimating Thresholds in Autoregressive Models, *Journal of Time Series Analysis*, 7, 179-190.

3. Chan, K.S. (1993), Consistency and Limiting Distribution of the Least Square Estimation of a Threshold Autoregressive Model, *Annals of Statistics,* 21, 520-533.

4. Chan, K.S. and R.S. Tsay (1998), Limiting Properties of the Least Square Estimation of a Continuous Threshold Autoregressive Model, *Biometrika,* 85, 2, 413-426.

5. Chan, K.S. (1990), Testing for Threshold Autoregression, *Annals of Statistics,* 18, 1886-1894.

6. Chen, R. and R.S. Tsay (1993), Functional-coefficient Autoregresives Models, *Journal of the American Statistical Association,* 88, 298-308.

7. Chong, T.L. and K.M. Yan (2004), Threshold Model with multiple threshold Variables and Application to Financial Crises, *Working paper.*

8. Chong, T.L. (2003), Generic Consistency of the Break-Point Estimator under Specification Errors, *Econometrics Journal,* 6, 167-192.

9. Dueker, M., S. Martin and F. Spagnolo (2003), Contemporaneous Threshold Autoregressive Models: Estimation, Forecasting and Rational Expectations Application, *Working paper.*

10. Enders, W. and C.W.J. Granger (1998), Unit Root Tests and Asymmetric Adjustment with an Example Using the Term Structure of Interest Rates, *Journal of Business and Economic Statistics,*16, 3, 304-311.

11. Fabozzi, F.J. and J.C. Francis (1977), Stability Tests for Alphas and Betas over Bull and Bear Market Conditions. *Journal of Finance,* XXX 11, *4, 1093-1099.*

12. Gonzalo, J., and J.Y. Pitarakis (2002), Estimation and Model Selection Based Inference in Single and Multiple Threshold Models, *Journal of Econometrics,* 110, 319-352.

13. Granville, J. (1963), *Granville's new key to stock market profits,* Prentice Hall.

14. Haggan, V. and T. Ozaki (1981), Modeling Vibrations Using a Amplitude-dependent Autoregressive Time Series Model, *Biometrika,* 68, 189-196.

15. Hansen, B.E. (2000), Sample Splitting and Threshold Estimation, *Econometrica,* 68, 3, 575-603.

16. Hansen, B.E. (1999), Threshold Effects in Non-dynamic Panels: Estimation, Testing and Inference, *Journal of Econometrics,* 93, 345-368.

17. Hansen, B.E. (1996), Inference when a Nuisance Parameter is not Identified under the Null Hypothesis, *Econometrica,* 64, 413-430.

18. Henry, Ó., N. Olekaln and P.M. Summers (2001), Exchange rate instability: a threshold autoregressive approach, *Economic Record* 77, 160-166.

19. Karpoff, J.M. (1987), The Relation between Price Changes and Trading Volume: A Survey" *Journal of Financial and Quantitative Analysis,* 22, 1,109-125.

20. Kim, M.K. and J.K. Zumwalt (1979), An Analysis of risk in bull and bear markets, *Journal of Financial and Quantitative Analysis,* XIV, 5,1015-1024.

21. Lee M.C. and B. Swaminathan (2000), Price Momentum and Trading Volume, *Journal of Finance,* IV, 5, 2017-2069.

22. Li, W.K. and K. Lam (1995), Modeling Asymmetry in Stock Returns by a Threshold Autoregressive Conditional Heteroscedastic Model, *Statistician* 44, 3, 333-341.

23. Lunde, A. and A. Timmermann (2004), Duration Dependence in Stock Prices: An Analysis of Bull and Bear Markets, *Journal of Business & Economic Statistics,* 22, 3, 253-279.

24. Lim, K.S. (1987), A comparative Study of Various Univariate Time Series Models for Canadian Lynx Data, *Journal of Time Series Analysis,* 8, 161-176.

25. Marktens, M., P. Kofman and C.F. Vorst (1998), A threshold error-correction model for intraday futures and index returns, *Journal of Applied Econometrics* 13, 245-263.

26. Potter, S.E. (1995), A nonlinear Approach to U.S. GNP, *Journal of Applied Econometrics,* 10, 109-125.

27. Petruccelli J., and N. Davies (1986), A Portmanteau Test for Self-exciting Threshold Autoregressive-Type Nonlinearity in Time Series, *Biometrika,* 73, 687-694.

28. Shively, A. (2003), The nonlinear dynamics of stock prices, *Quarterly Review of Economics and Finance* 43, 505-517.

29. Tsay, R.S. (1989), Testing and Modeling Threshold Auotregressive process, *Journal of the American Statistical Association,* 84, 405,1188-1201.

30. Tsay, R.S. (1998), Testing and Modeling Multivariate Threshold Models, *Journal of the American Statistical Association,* 93, 443, 231-240.

31. Tong, H. (1978), *On a Threshold Model in a Pattern Recognition and Signal Processing,* ed. C.H. Chen, Ameterdam: Sijhoff & Noordhoff.

32. Tong, H., and K.S. Lim (1980), Threshold Autoregression, Limit Cycles and Cyclical Data, *Journal of The Royal Statistical Society,* Ser.B, 42, 245-292.

33. Tong, H. (1983), *Threshold models in Non-linear Time Series Analysis* (Lecture Notes in Statistics No. 21), New York: Springer-Verlag.

APPENDIX 1: Proof of Theorem 1

The true threshold model is:

$$Y = \sum_{j=1}^{4} I_j(\gamma^0) X\beta_j + U.$$

The estimated model is:

$$Y = \sum_{j=1}^{4} I_j(\gamma) X\widehat{\beta}_j + \widehat{U}.$$

where

$$I_j(\gamma) = diag\left\{\Psi_T^{(j)}(\gamma), \Psi_{T-1}^{(j)}(\gamma), ..., \Psi_{p+1}^{(j)}(\gamma)\right\}$$

and

$\Psi_t^{(1)}(\gamma) = I(z_{1t} \le \gamma_1, z_{2t} \le \gamma_2);$

$\Psi_t^{(2)}(\gamma) = I(z_{1t} \le \gamma_1, z_{2t} > \gamma_2);$

$\Psi_t^{(3)}(\gamma) = I(z_{1t} > \gamma_1, z_{2t} \le \gamma_2);$

$\Psi_t^{(4)}(\gamma) = I(z_{1t} > \gamma_1, z_{2t} > \gamma_2).$

The residual sum of squares is defined as:

$\frac{1}{T}RSS_T(\gamma) = \frac{1}{T}|| \sum_{i=1}^{4} I_i(\gamma^0)X\beta_i + U - \sum_{i=1}^4 I_i(\gamma)X\widehat{\beta}_i||^2$

$= \frac{1}{T}\left[\sum_{i=1}^{4}\left(\beta_i' X' I_i(\gamma^0)X\beta_i + \widehat{\beta}_i' X' I_i(\gamma)X\widehat{\beta}_i\right) - 2\sum_{i=1}^4\sum_{j=1}^4 \beta_i' X' I_i(\gamma^0)I_j(\gamma)X\widehat{\beta}_j\right]$

$+\frac{2}{T}U'\sum_{i=1}^{4}\left(I_i(\gamma^0)X\beta_i - I_i(\gamma)X\widehat{\beta}_i\right) + \frac{1}{T}U'U.$

For any γ, we can define the following moment functionals:

$$M_i(\gamma) = X' I_i(\gamma)X, \; i = 1, 2, 3, 4.$$

Next, we prove that $RSS_T(\gamma)$ is minimized if and only if $\gamma = \gamma^0$ by partitioning the threshold space into four regions.

Case 1: $\gamma_1 \le \gamma_1^0$, and $\gamma_2 \le \gamma_2^0$

$I_1(\gamma)I_1(\gamma^0) = I_1(\gamma),$

$I_2(\gamma)I_1(\gamma^0) = I_1(\gamma_1, \gamma_2^0) - I_1(\gamma),$

$I_3(\gamma)I_1(\gamma^0) = I_1(\gamma_1^0, \gamma_2) - I_1(\gamma),$

$I_4(\gamma)I_1(\gamma^0) = I_1(\gamma^0) + I_1(\gamma) - I_1(\gamma_1^0, \gamma_2) - I_1(\gamma_1, \gamma_2^0)$.

We have

$\widehat{\beta}_1(\gamma) = (X'I_1(\gamma)X)^{-1}X'I_1(\gamma)Y = (X'I_1(\gamma)X)^{-1}X'I_1(\gamma)[\sum_{i=1}^{4} I_i(\gamma^0)X\beta_i + U]$

$\xrightarrow{p} \beta_1 = (\beta_1 - \beta_2) + \beta_2;$

$\widehat{\beta}_2(\gamma) = (X'I_2(\gamma)X)^{-1}X'I_2(\gamma)Y = (X'I_2(\gamma)X)^{-1}X'I_2(\gamma)[\sum_{i=1}^{4} I_i(\gamma^0)X\beta_i + U]$

$\xrightarrow{p} \dfrac{M_2(\gamma) - M_2(\gamma_1, \gamma_2^0)}{M_2(\gamma)}(\beta_1 - \beta_2) + \beta_2;$

$\widehat{\beta}_3(\gamma) = (X'I_3(\gamma)X)^{-1}X'I_3(\gamma)Y = (X'I_3(\gamma)X)^{-1}X'I_3(\gamma)\left[\sum_{i=1}^{4} I_i(\gamma^0)X\beta_i + U\right]$

$\xrightarrow{p} \dfrac{M_3(\gamma) - M_3(\gamma_1^0, \gamma_2)}{M_3(\gamma)}(\beta_1 - \beta_2) + \dfrac{M_3(\gamma_1^0, \gamma_2)}{M_3(\gamma)}(\beta_3 - \beta_2) + \beta_2;$

$\widehat{\beta}_4(\gamma) = (X'I_4(\gamma)X)^{-1}X'I_4(\gamma)Y$

$= (X'I_4(\gamma)X)^{-1}X'I_4(\gamma)[I_1(\gamma^0)X\beta_1 + (I - I_1(\gamma^0))X\beta_2 + U]$

$\xrightarrow{p} \dfrac{M_4(\gamma^0) - M_4(\gamma_1^0, \gamma_2) - M_4(\gamma_1, \gamma_2^0) + M_4(\gamma)}{M_4(\gamma)}(\beta_1 - \beta_2) + \beta_2 +$

$\dfrac{M_4(\gamma^0)}{M_4(\gamma)}(\beta_4 - \beta_2) + \dfrac{M_4(\gamma_1^0, \gamma_2) - M_4(\gamma^0)}{M_4(\gamma)}(\beta_3 - \beta_2)$

and we can get

$\frac{1}{T}RSS_T(\gamma) = \sum_{i=1}^{4}\left(\beta_i' M_i(\gamma^0)\beta_i + \widehat{\beta}_i' M_i(\gamma)\widehat{\beta}_i\right) -$

$2\sum_{i=1}^{4}\sum_{j=1}^{4} \beta_i' X' I_i(\gamma^0) I_j(\gamma) X\widehat{\beta}_j + \sigma^2 + o_p(1)$

$= \sum_{i=1}^{4} \beta_i' M_i(\gamma^0)\beta_i - \sum_{i=1}^{4} \widehat{\beta}_i' M_i(\gamma)\beta_2 -$

$$\begin{bmatrix} M_1(\gamma)\widehat{\beta}_1 + (M_2(\gamma) - M_2(\gamma_1, \gamma_2^0))\widehat{\beta}_2 + [M_3(\gamma) - M_3(\gamma_1^0, \gamma_2)]\widehat{\beta}_3 \\ +[M_4(\gamma^0) - M_4(\gamma_1^0, \gamma_2) - M_4(\gamma_1, \gamma_2^0) + M_4(\gamma)]\widehat{\beta}_4 \end{bmatrix}'(\beta_1 - \beta_2)$$

$-\left[M_3(\gamma_1^0, \gamma_2)\widehat{\beta}_3(\gamma) + (M_4(\gamma_1^0, \gamma_2) - M_4(\gamma^0))\widehat{\beta}_4(\gamma)\right]'(\beta_3 - \beta_2) -$

$\widehat{\beta}_4(\gamma)' M_4(\gamma^0)(\beta_4 - \beta_2) + \sigma^2 + o_p(1)$

$= \sum_{i=1}^{4} \beta_i' M_i(\gamma^0)(\beta_i - \beta_2) -$

$$\begin{bmatrix} M_1(\gamma)\widehat{\beta}_1 + (M_2(\gamma) - M_2(\gamma_1, \gamma_2^0))\widehat{\beta}_2 + [M_3(\gamma) - M_3(\gamma_1^0, \gamma_2)]\widehat{\beta}_3 + \\ [M_4(\gamma^0) - M_4(\gamma_1^0, \gamma_2) - M_4(\gamma_1, \gamma_2^0) + M_4(\gamma)]\widehat{\beta}_4 \end{bmatrix}'(\beta_1 - \beta_2)$$

$-\left[M_3(\gamma_1^0, \gamma_2)\widehat{\beta}_3(\gamma) + (M_4(\gamma_1^0, \gamma_2) - M_4(\gamma^0))\widehat{\beta}_4(\gamma)\right]'(\beta_3 - \beta_2) -$

$\left[M_4(\gamma^0)\widehat{\beta}_4(\gamma)\right]'(\beta_4 - \beta_2) + \sigma^2 + o_p(1)$

$$= (\beta_1 - \beta_2)'\begin{bmatrix} M_1(\gamma^0) - M_1(\gamma) - \dfrac{(M_2(\gamma) - M_2(\gamma_1, \gamma_2^0))^2}{M_2(\gamma)} - \dfrac{(M_3(\gamma) - M_3(\gamma_1^0, \gamma_2))^2}{M_3(\gamma)} \\ -\dfrac{(M_4(\gamma^0) - M_4(\gamma_1^0, \gamma_2) - M_4(\gamma_1, \gamma_2^0) + M_4(\gamma))^2}{M_4(\gamma)} \end{bmatrix}(\beta_1 -$$

$\beta_2)$

$+(\beta_3-\beta_2)'\left[M_3(\gamma^0)-\frac{(M_4(\gamma_1^0,\gamma_2)-M_4(\gamma^0))^2}{M_4(\gamma)}-\frac{(M_3(\gamma_1^0,\gamma_2))^2}{M_3(\gamma)}\right](\beta_3-\beta_2)+$

$(\beta_4-\beta_2)'\left[M_4(\gamma^0)-\frac{(M_4(\gamma^0))^2}{M_4(\gamma)}\right](\beta_4-\beta_2)+o_p(1)=(\beta_1-\beta_2)'Q_1(\beta_1-\beta_2)$

$+(\beta_3-\beta_2)'Q_2(\beta_3-\beta_2)+(\beta_4-\beta_2)'Q_3(\beta_4-\beta_2)+o_p(1)=b_1(\gamma)+\sigma^2+o_p(1)$

for any $\gamma_1<\gamma_1^0$, and $\gamma_2<\gamma_2^0$, using following results:

$I_2(\gamma)-I_2(\gamma_1,\gamma_2^0)<I_2(\gamma)$,

$I_3(\gamma)-I_3(\gamma_1^0,\gamma_2)<I_3(\gamma)$,

$I_4(\gamma^0)-I_4(\gamma_1^0,\gamma_2)-I_4(\gamma_1,\gamma_2^0)+I_4(\gamma)<I_4(\gamma)$,

$I_4(\gamma^0)<I_4(\gamma)$,

$I_4(\gamma_1^0,\gamma_2)-I_4(\gamma^0)<I_4(\gamma)$,

$I_3(\gamma_1^0,\gamma_2)<I_3(\gamma)$,

we can prove that Q_1, Q_2 and Q_3 are positive definite matrices, thus.$b_1(\gamma)\geq b_1(\gamma^0)=0$, and the equation holds if and only if $\gamma=\gamma^0$.

Case 2: $\gamma_1>\gamma_1^0$, and $\gamma_2<\gamma_2^0$, we have

$\widehat{\beta}_1(\gamma)=(X'I_1(\gamma)X)^{-1}X'I_1(\gamma)Y=(X'I_1(\gamma)X)^{-1}X'I_1(\gamma)[\sum_{i=1}^{4}I_i(\gamma^0)X\beta_i+U]$

$\xrightarrow{p}\frac{M_1(\gamma)-M_1(\gamma_1^0,\gamma_2)}{M_1(\gamma)}(\beta_3-\beta_1)+\beta_1;$

$\widehat{\beta}_2(\gamma)=(X'I_2(\gamma)X)^{-1}X'I_2(\gamma)Y=(X'I_2(\gamma)X)^{-1}X'I_2(\gamma)[\sum_{i=1}^{4}I_i(\gamma^0)X\beta_i+U]$

$\xrightarrow{p}\frac{M_2(\gamma^0)-M_2(\gamma_1^0,\gamma_2)-M_2(\gamma_1,\gamma_2^0)+M_2(\gamma)}{M_2(\gamma)}(\beta_3-\beta_1)+\beta_1+\frac{M_2(\gamma^0)}{M_2(\gamma)}(\beta_2-\beta_1)+$

$\frac{M_2(\gamma_1,\gamma_2^0)-M_2(\gamma^0)}{M_2(\gamma)}(\beta_4-\beta_1);$

$\widehat{\beta}_3(\gamma)=(X'I_3(\gamma)X)^{-1}X'I_3(\gamma)Y=(X'I_3(\gamma)X)^{-1}X'I_3(\gamma)[\sum_{i=1}^{4}I_i(\gamma^0)X\beta_i+U]$

$\xrightarrow{p}(\beta_3-\beta_1)+\beta_1;$

$\widehat{\beta}_4(\gamma)=(X'I_4(\gamma)X)^{-1}X'I_4(\gamma)Y=(X'I_4(\gamma)X)^{-1}X'I_4(\gamma)[I_1(\gamma^0)X\beta_1+(I-I_1(\gamma^0))X\beta_2+U]$

$\xrightarrow{p}\frac{M_4(\gamma_1,\gamma_2^0)}{M_4(\gamma)}(\beta_4-\beta_1)+\frac{M_4(\gamma)-M_4(\gamma_1,\gamma_2^0)}{M_4(\gamma)}(\beta_3-\beta_1)+\beta_1;$

and

$\frac{1}{T}RSS_T(\gamma)=\sigma^2+\sum_{i=1}^{4}\left(\beta_i'M_i(\gamma^0)\beta_i+\widehat{\beta}_i'M_i(\gamma)\widehat{\beta}_i\right)-$
$2\sum_{i=1}^{4}\sum_{j=1}^{4}\beta_i'X'I_i(\gamma^0)I_j(\gamma)X\widehat{\beta}_j+o_p(1)$

$= \sum_{i=1}^{4} \beta_i' M_i(\gamma^0)\beta_i - \sum_{i=1}^{4} \widehat{\beta}_i' M_i(\gamma)\beta_1 -$

$$\begin{bmatrix} M_3(\gamma)\widehat{\beta}_3 + (M_1(\gamma) - M_1(\gamma_1^0,\gamma_2))\widehat{\beta}_1 + [M_4(\gamma) - M_4(\gamma_1,\gamma_2^0)]\widehat{\beta}_4 + \\ [M_2(\gamma^0) - M_2(\gamma_1^0,\gamma_2) - M_2(\gamma_1,\gamma_2^0) + M_2(\gamma)]\widehat{\beta}_2 \end{bmatrix}' (\beta_3 - \beta_1)$$

$- \left[M_4(\gamma_1,\gamma_2^0)\widehat{\beta}_4(\gamma) + (M_2(\gamma_1,\gamma_2^0) - M_2(\gamma^0))\widehat{\beta}_2(\gamma)\right]' (\beta_4 - \beta_1) - \widehat{\beta}_2(\gamma)' M_2(\gamma^0)(\beta_2 -$
$\beta_1) + \sigma^2 + o_p(1)$

$$= (\beta_3 - \beta_1)' \begin{bmatrix} M_3(\gamma^0) - M_3(\gamma) - \dfrac{(M_1(\gamma) - M_1(\gamma_1^0,\gamma_2))^2}{M_1(\gamma)} - \dfrac{(M_4(\gamma) - M_4(\gamma_1,\gamma_2^0))^2}{M_4(\gamma)} \\ - \dfrac{(M_2(\gamma^0) - M_2(\gamma_1^0,\gamma_2) - M_2(\gamma_1,\gamma_2^0) + M_2(\gamma))^2}{M_2(\gamma)} \end{bmatrix} (\beta_3 -$$

$\beta_1)$

$+(\beta_4 - \beta_1)' \left[(M_4(\gamma^0) - \dfrac{(M_4(\gamma_1,\gamma_2^0))^2}{M_4(\gamma)} - \dfrac{(M_2(\gamma_1,\gamma_2^0) - M_2(\gamma^0))^2}{M_2(\gamma)}\right] (\beta_4 - \beta_1) +$

$(\beta_2 - \beta_1)' \left[M_2(\gamma^0) - \dfrac{(M_2(\gamma^0))^2}{M_2(\gamma)}\right] (\beta_2 - \beta_1) + \sigma^2 + o_p(1) = (\beta_3 - \beta_1)' Q_4 (\beta_3 - \beta_1)$

$+(\beta_4 - \beta_1)' Q_5 (\beta_4 - \beta_1) + (\beta_2 - \beta_1)' Q_6 (\beta_2 - \beta_1) + \sigma^2 + o_p(1) = b_2(\gamma) + \sigma^2 + o_p(1).$

for any $\gamma_1 > \gamma_1^0$, and $\gamma_2 < \gamma_2^0$, using following results:

$I_2(\gamma^0) - I_2(\gamma_1^0,\gamma_2) - I_2(\gamma_1,\gamma_2^0) + I_2(\gamma) < I_2(\gamma)$,

$I_2(\gamma^0) < I_2(\gamma)$,

$I_2(\gamma_1,\gamma_2^0) - I_2(\gamma^0) < I_4(\gamma)$,

$I_4(\gamma_1,\gamma_2^0) < I_4(\gamma)$,

we can prove that Q_4, Q_5 and Q_6 are positive definite matrices, thus $b_2(\gamma) > b_2(\gamma^0) = 0$ for any $\gamma_1 > \gamma_1^0$, and $\gamma_2 \leq \gamma_2^0$.

Since Case 3 and Case 4 are mirror images of the Case 2 and Case 1, we can use similar method to prove that

$\frac{1}{T} RSS_T(\gamma) - \sigma^2 = b_i(\gamma) + o_p(1)$ and $b_i(\gamma) \geq b_i(\gamma^0) = 0$ in both cases.

Define a non-stochastic function $b(\gamma)$ as $b_i(\gamma)$ in the *ith* case, then

$$\sup_{\gamma \in R^2} \left|\frac{1}{T} RSS_T(\gamma) - b(\gamma) - \sigma^2\right| = o_p(1)$$

$b(\gamma)$ is minimized if and only if $\gamma = \gamma^0$, so $RSS_T(\gamma)$ is minimized at the true thresholds and $\widehat{\gamma}$ are consistent.

The proof for the consistency of the structural parameters $\widehat{\beta}_i(\widehat{\gamma})$ is easily obtained.

APPENDIX 2: Proof of Theorem 2

To derive the limiting distribution of $\widehat{\gamma}$ for shrinking break, we let $\delta = cT^{-\alpha}$, $0 < \alpha < \frac{1}{2}$, c is a constant vector.

$$\widehat{\gamma} = \arg\min_{\gamma\in\Omega} RSS_T(\gamma) = \arg\min_{\gamma\in\Omega} \left[RSS_T(\gamma) - RSS_T\left(\gamma^0\right)\right].$$

where $\Omega = [\underline{\gamma_1}, \overline{\gamma_1}] \times [\underline{\gamma_2}, \overline{\gamma_2}]$

The model is:

$$Y = \sum_{j=2}^{4} X_{\gamma^0}^{(j)}\beta_1 + U.$$

The model can be rewritten as:

$$Y = X\beta_1 + \sum_{j=2}^{4} X_{\gamma^0}^{(j)}\delta^{(j)} + U.$$

Define $X_0 = \sum_{j=2}^{4} X_0^{(j)}$, and $X_0^{(j)} = X_{\gamma^0}^{(j)}$.

For any γ, define $X_\gamma = \sum_{j=2}^{4} X_\gamma^{(j)}$

Let $\widehat{\beta}_1 = \widehat{\beta}_1(\gamma)$, $\widehat{\beta}_1^0 = \widehat{\beta}_1(\gamma^0)$, $\widehat{\boldsymbol{\delta}}^{(j)} = \widehat{\boldsymbol{\delta}}^{(j)}(\gamma)$ and $\widehat{\boldsymbol{\delta}}_0^{(j)} = \widehat{\boldsymbol{\delta}}^{(j)}(\gamma^0)$.

In the neighborhood of the true thresholds, where $\gamma_1 = \gamma_1^0 + \frac{v}{T^{1-2\alpha}}$, $\gamma_2 = \gamma_2^0 + \frac{\omega}{T^{1-2\alpha}}$, we have

$\widehat{\beta}_1 = (X_\gamma^{(1)\prime}X_\gamma^{(1)})^{-1}X_\gamma^{(1)\prime}Y = \left(X'X - X_\gamma'X_\gamma\right)^{-1}(X - X_\gamma)'Y'$

$= \beta_1 + \sum_{j=2}^{4}\left(X'X - X_\gamma'X_\gamma\right)^{-1}\left(X_0^{(j)\prime}X_0^{(j)} - X_\gamma^{(j)\prime}X_0^{(j)}\right)'\delta^{(j)} +$

$\left(X'X - X_\gamma'X_\gamma\right)^{-1}(X - X_\gamma)'\varepsilon.$

When γ equals the true value,

$\widehat{\beta}_1(\gamma^0) = (X'X - X_0'X_0)^{-1}(X - X_0)'Y'$

$= \beta_1 + (X'X - X_0'X_0)^{-1}(X - X_0)'U.$

Then

$\widehat{\beta}_1(\gamma) - \widehat{\beta}_1(\gamma^0) = \sum_{j=2}^{4}\left(X'X - X_\gamma'X_\gamma\right)^{-1}\left(X_0^{(j)\prime}X_0^{(j)} - X_\gamma^{(j)\prime}X_0^{(j)}\right)'\delta^{(j)}$

$+ \left(X'X - X_\gamma'X_\gamma\right)^{-1}(X - X_\gamma)'U - (X'X - X_0'X_0)^{-1}(X - X_0)'U$

$= \sum_{j=2}^{4}\left(X'X - X_\gamma'X_\gamma\right)^{-1}\left(X_0^{(j)\prime}X_0^{(j)} - X_\gamma^{(j)\prime}X_0^{(j)}\right)'\delta^{(j)} + \left(X'X - X_\gamma'X_\gamma\right)^{-1}(X_0 - X_\gamma)'U +$

$O_p\left(\frac{1}{T^{1-\alpha}}\right)$

$= O_p\left(\frac{1}{T^{1-\alpha}}\right)$,

and

$\widehat{\delta}^{(j)}(\gamma) = \left(X_\gamma^{(j)\prime} X_\gamma^{(j)}\right)^{-1} X_\gamma^{(j)\prime} Y - \beta_1 = \left(X_\gamma' X_\gamma\right)^{-1} X_\gamma'(X\beta_1 + X_0\delta + U)$

$= \left(X_\gamma' X_\gamma\right)^{-1} X_\gamma' X_0 \delta + \left(X_\gamma' X_\gamma\right)^{-1} X_\gamma' U$

$= O_p\left(\frac{1}{T^{1-\alpha}}\right)$.

Thus,

$$\widehat{\delta}^{(j)} - \widehat{\delta}_0^{(j)} = O_p\left(\frac{1}{T^{1-\alpha}}\right).$$

We also have

$$\widehat{\delta}_0^{(j)} - \delta^{(j)} = O_p\left(T^{-\frac{1}{2}}\right).$$

$$RSS_T(\gamma) - RSS_T(\gamma^0) = \left(Y - X\widehat{\beta}_1 - \sum_{j=2}^4 X_\gamma^{(j)} \widehat{\boldsymbol{\delta}}^{(j)}\right)' \left(Y - X\widehat{\beta}_1 - \sum_{j=2}^4 X_\gamma^{(j)} \widehat{\boldsymbol{\delta}}^{(j)}\right) - RSS_T(\gamma^0)$$

$$= Y'Y - 2Y'\left(X\widehat{\beta}_1 + \sum_{j=2}^4 X_\gamma^{(j)} \widehat{\boldsymbol{\delta}}^{(j)}\right) + \left(X\widehat{\beta}_1 + \sum_{j=2}^4 X_\gamma^{(j)} \widehat{\boldsymbol{\delta}}^{(j)}\right)' \left(X\widehat{\beta}_1 + \sum_{j=2}^4 X_\gamma^{(j)} \widehat{\boldsymbol{\delta}}^{(j)}\right) - RSS_T(\gamma^0)$$

$$= -2\left(X\beta_1 + \sum_{j=2}^4 X_0^{(j)} \delta^{(j)} + U\right)' \left(X\left(\widehat{\beta}_1 - \widehat{\beta}_1^0\right) + \sum_{j=2}^4 X_\gamma^{(j)} \widehat{\boldsymbol{\delta}}^{(j)} - \sum_{j=2}^4 X_0^{(j)} \widehat{\boldsymbol{\delta}}_0^{(j)}\right)$$

$$+ \left(X\widehat{\beta}_1 + \sum_{j=2}^4 X_\gamma^{(j)} \widehat{\boldsymbol{\delta}}^{(j)}\right)' \left(X\widehat{\beta}_1 + \sum_{j=2}^4 X_\gamma^{(j)} \widehat{\boldsymbol{\delta}}^{(j)}\right)$$

$$- \left(X\widehat{\beta}_1^0 + \sum_{j=2}^4 X_0^{(j)} \widehat{\boldsymbol{\delta}}_0^{(j)}\right)' \left(X\widehat{\beta}_1^0 + \sum_{j=2}^4 X_0^{(j)} \widehat{\boldsymbol{\delta}}_0^{(j)}\right)$$

$$= -2\left(X\beta_1 + \sum_{j=2}^4 X_0^{(j)} \delta^{(j)} + U\right)' \left(\sum_{j=2}^4 X_\gamma^{(j)} \widehat{\boldsymbol{\delta}}^{(j)} - \sum_{j=2}^4 X_0^{(j)} \widehat{\boldsymbol{\delta}}_0^{(j)}\right) +$$

$$\left(X\widehat{\beta}_1 + \sum_{j=2}^4 X_\gamma^{(j)} \widehat{\boldsymbol{\delta}}^{(j)} + X\widehat{\beta}_1^0 + \sum_{j=2}^4 X_0^{(j)} \widehat{\boldsymbol{\delta}}_0^{(j)}\right)' \left(X\widehat{\beta}_1 + \sum_{j=2}^4 X_\gamma^{(j)} \widehat{\boldsymbol{\delta}}^{(j)} - X\widehat{\beta}_1^0 - \sum_{j=2}^4 X_0^{(j)} \widehat{\boldsymbol{\delta}}_0^{(j)}\right) + o_p(1)$$

$$= -2U'\left(\sum_{j=2}^4 X_\gamma^{(j)} \widehat{\boldsymbol{\delta}}^{(j)} - \sum_{j=2}^4 X_0^{(j)} \widehat{\boldsymbol{\delta}}_0^{(j)}\right)$$

$$+ \left(\sum_{j=2}^4 X_\gamma^{(j)} \widehat{\boldsymbol{\delta}}^{(j)} - \sum_{j=2}^4 X_0^{(j)} \widehat{\boldsymbol{\delta}}_0^{(j)}\right)' \left(\sum_{j=2}^4 X_\gamma^{(j)} \widehat{\boldsymbol{\delta}}^{(j)} - \sum_{j=2}^4 X_0^{(j)} \widehat{\boldsymbol{\delta}}_0^{(j)}\right) + o_p(1)$$

Define $X_\gamma^* = \left(X_\gamma^{(2)}, X_\gamma^{(3)}, X_\gamma^{(4)}\right)$ and $\delta = (\boldsymbol{\delta}^{(2)}, \boldsymbol{\delta}^{(3)}, \boldsymbol{\delta}^{(4)}) = cT^{-\alpha}, 0 < \alpha < \frac{1}{2}$, c is a constant vector, then

$RSS_T(\gamma) - RSS_T(\gamma^0) = -2U'(X_\gamma^* - X_0^*)\delta + \left((X_\gamma^* - X_0^*)\delta\right)' \left((X_\gamma^* - X_0^*)\delta\right) + o_p(1)$
$= -2U'(X_\gamma^* - X_0^*)\delta + \delta' \left(X_\gamma^* - X_0^*\right)' \left(X_\gamma^* - X_0^*\right) \delta^{(j)} + o_p(1)$
$= -2U'(X_\gamma^* - X_0^*)cT^{-\alpha} + c' \left(X_\gamma^* - X_0^*\right)' \left(X_\gamma^* - X_0^*\right) cT^{-2\alpha} + o_p(1).$

Now, using the expansion of the moment functionals:

$$M_i(\gamma) = M_i\left(\gamma^0\right) + \left(\gamma_1 - \gamma_1^0\right) DF_{i1}^0 + \left(\gamma_2 - \gamma_2^0\right) DF_{i2}^0 + o(1),$$

$$M_i\left(\gamma_1^0, \gamma_2\right) = M_i\left(\gamma^0\right) + \left(\gamma_2 - \gamma_2^0\right) DF_{i2}^0 + o(1),$$

$$M_i\left(\gamma_1, \gamma_2^0\right) = M_i\left(\gamma^0\right) + \left(\gamma_1 - \gamma_1^0\right) DF_{i1}^0 + o(1),$$

where, $F_{i1}^0 = \frac{\partial F_i(\gamma)}{\partial \gamma_1}|_{\gamma=\gamma^0}$, $F_{i2}^0 = \frac{\partial F_i(\gamma)}{\partial \gamma_2}|_{\gamma=\gamma^0}$,

and from the assumption that $f(\gamma)$ is continuous at the $\gamma = \gamma^0$, we have

$F_{11}^0 = F_{21}^0 = -F_{31}^0 = -F_{41}^0 = F_1^0 = \frac{\partial F(\gamma)}{\partial(\gamma_1)}|_{\gamma=\gamma^0}$,

$F_{12}^0 = F_{32}^0 = -F_{22}^0 = -F_{42}^0 = F_2^0 = \frac{\partial F(\gamma)}{\partial(\gamma_2)}|_{\gamma=\gamma^0}$,

where $F(\gamma)$ is the joint distribution function and $f(\gamma)$ is the joint density function.

Let $D^* = (D', D', D')'$ and $V^* = (V', V', V')'$.

We can prove that

$$\left(X_\gamma^* - X_0^*\right)' \left(X_\gamma^* - X_0^*\right) \xrightarrow{p} |\gamma_1 - \gamma_1^0| D^* F_1^0 + |\gamma_2 - \gamma_2^0| D^* F_2^0,$$

$RSS_T(\gamma) - RSS_T(\gamma^0) = -2U'(X_\gamma^* - X_0^*)cT^{-\alpha} + c' \left(X_\gamma^* - X_0^*\right)' \left(X_\gamma^* - X_0^*\right) cT^{-2\alpha} + o_p(1)$

$\overset{d}{=} c' D^* c |\upsilon| F_1^0 - 2c' B_1(\upsilon) + c' D^* c |\omega| F_2^0 - 2c' B_2(\omega)$,

where $B_1(\upsilon)$ is a vector Brownian motion with covariance matrix $E\left(B_1(1) B_1(1)'\right) = V^* F_1^0$,

and $B_2(\omega)$ is a vector Brownian motion with covariance matrix $E\left(B_2(1) B_2(1)'\right) = V^* F_2^0$.

Make the change-of-variables

$$\upsilon = \frac{c' V^* c}{(c' D^* c)^2 F_1^0} r_1,$$

$$\omega = \frac{c'V^*c}{\left(c'D^*c\right)^2 F_2^0} r_2.$$

Thus,

$$\begin{aligned}
&RSS_T\left(\gamma\right) - RSS_T\left(\gamma^0\right) \\
&\overset{d}{=} \frac{c'V^*c}{c'D^*c}\left|r_1\right| - 2\frac{\sqrt{c'V^*c}}{c'D^*c\sqrt{F_1^0}}{}' B_1\left(v\right) + \frac{c'V^*c}{c'D^*c}\left|r_2\right| - 2\frac{\sqrt{c'V^*c}}{c'D^*c\sqrt{F_2^0}} c' B_2\left(w\right) \\
&\overset{d}{=} \frac{c'V^*c}{c'D^*c}\left|r_1\right| - 2\frac{c'V^*c}{c'D^*c} W_1\left(-r_1\right) + \frac{c'V^*c}{c'D^*c}\left|r_2\right| - 2\frac{c'V^*c}{c'D^*c} W_2\left(-r_2\right) \\
&\overset{d}{=} 2\frac{c'V^*c}{c'D^*c}\left(\frac{\left|r_1\right|}{2} + W_1\left(r_1\right) + \frac{\left|r_2\right|}{2} + W_2\left(r_2\right)\right).
\end{aligned}$$

We have

$$\begin{aligned}
& T^{1-2\alpha}\frac{\left(c'D^*c\right)^2}{c'V^*c}\Big((\widehat{\gamma}_1 - \gamma_1^0)F_1^0, (\widehat{\gamma}_2 - \gamma_2^0)F_2^0\Big) \\
= \;& \left(\widehat{r}_1, \widehat{r}_2\right) \\
& \overset{d}{\to} \arg \min_{-\infty < r_1 < \infty, -\infty < r_2 < \infty} \left(\frac{\left|r_1\right|}{2} + W_1\left(r_1\right) + \frac{\left|r_2\right|}{2} + W_2\left(r_2\right)\right) \\
& \overset{d}{=} \arg \max_{-\infty < r_1 < \infty, -\infty < r_2 < \infty} \left(-\frac{1}{2}\left|r_1\right| + W_1\left(r_1\right) - \frac{1}{2}\left|r_2\right| + W_2\left(r_2\right)\right).
\end{aligned}$$

APPENDIX 3: Proof of Theorem 3

We first consider the case where the number of regimes is misspecified. Suppose the true model is

$$Y = I_1(\gamma^0)F\beta_1 + [I - I_1(\gamma^0)]F\beta_2 + U.$$

However, the estimated model misspecifies the number of regimes as 4.

$$Y = \sum_{i=1}^{4} I_i(\gamma)F\widehat{\beta}_i + \widehat{U},$$

where,

$$I_j(\gamma) = diag\left\{\Psi_T^{(j)}\left(\gamma\right), \Psi_{T-1}^{(j)}\left(\gamma\right), ..., \Psi_{T-p}^{(j)}\left(\gamma\right)\right\}$$

and

$$\Psi_t^{(1)}(\gamma) = I(z_{1t} \leq \gamma_1, z_{2t} \leq \gamma_2)$$
$$\Psi_t^{(2)}(\gamma) = I(z_{1t} \leq \gamma_1, z_{2t} > \gamma_2)$$
$$\Psi_t^{(3)}(\gamma) = I(z_{1t} > \gamma_1, z_{2t} \leq \gamma_2)$$
$$\Psi_t^{(4)}(\gamma) = I(z_{1t} > \gamma_1, z_{2t} > \gamma_2).$$

We have the residual sum of squares as:

$\frac{1}{T}RSS_T(\gamma) = \frac{1}{T}||I_1(\gamma^0)F\beta_1 + [I - I_1(\gamma^0)]F\beta_2 + U - \sum_{i=1}^4 I_i(\gamma)F\widehat{\beta}_i||^2$
$= \frac{1}{T}[U'U + \beta_1'F'I_1(\gamma^0)F\beta_1 + \beta_2'F'[I - I_1(\gamma^0)]F\beta_2 + \sum_{i=1}^4 \widehat{\beta}_i'F'I_i(\gamma)F\widehat{\beta}_i$
$-2\sum_{i=1}^4 \beta_1'F'I_1(\gamma^0)I_i(\gamma)F\widehat{\beta}_i - 2\sum_{i=1}^4 \beta_2'F'[I - I_1(\gamma^0)]I_i(\gamma)F\widehat{\beta}_i]$
$+\frac{2}{T}U'[I_1(\gamma^0)F\beta_1 + [I - I_1(\gamma^0)]F\beta_2 - \sum_{i=1}^4 I_i(\gamma)F\widehat{\beta}_i]$
$= \frac{1}{T}[U'U + \beta_1'F'I_1(\gamma^0)F\beta_1 + \beta_2'F'[I - I_1(\gamma^0)]F\beta_2 - \sum_{i=1}^4 \widehat{\beta}_i'F'I_i(\gamma)F\widehat{\beta}_i +$
$\frac{2}{T}U'[I_1(\gamma^0)F\beta_1 + [I - I_1(\gamma^0)]F\beta_2 - \sum_{i=1}^4 I_i(\gamma)F\widehat{\beta}_i].$

Let $Q_{ff} = E(F'F)$.

For any γ, we can define the following moment functionals:

$$M_i(\gamma) = F'I_i(\gamma)F, \; i = 1, 2, 3, 4.$$

Next, we prove that $RSS_T(\gamma)$ is minimized at $\gamma = \gamma^0$ uniquely.

Case 1:$\gamma_1 \leq \gamma_1^0$, and $\gamma_2 \leq \gamma_2^0$

$I_1(\gamma)I_1(\gamma^0) = I_1(\gamma),$
$I_2(\gamma)I_1(\gamma^0) = I_1(\gamma_1, \gamma_2^0) - I_1(\gamma),$
$I_3(\gamma)I_1(\gamma^0) = I_1(\gamma_1^0, \gamma_2) - I_1(\gamma),$
$I_4(\gamma)I_1(\gamma^0) = I_1(\gamma^0) + I_1(\gamma) - I_1(\gamma_1^0, \gamma_2) - I_1(\gamma_1, \gamma_2^0).$

We have

$\widehat{\beta}_1(\gamma) = (F'I_1(\gamma)F)^{-1}F'I_1(\gamma)Y$
$= (F'I_1(\gamma)F)^{-1}F'I_1(\gamma)[I_1(\gamma^0)F\beta_1 + (I - I_1(\gamma^0))F\beta_2 + U]$
$\xrightarrow{p} \beta_1 = (\beta_1 - \beta_2) + \beta_2;$
$\widehat{\beta}_2(\gamma) = (F'I_2(\gamma)F)^{-1}F'I_2(\gamma)Y$
$= (F'I_2(\gamma)F)^{-1}F'I_2(\gamma)[I_1(\gamma^0)F\beta_1 + (I - I_1(\gamma^0))F\beta_2 + U]$
$\xrightarrow{p} \dfrac{M_1(\gamma_1, \gamma_2^0) - M_1(\gamma)}{M_2(\gamma)}(\beta_1 - \beta_2) + \beta_2;$
$\widehat{\beta}_3(\gamma) = (F'I_3(\gamma)F)^{-1}F'I_3(\gamma)Y$
$= (F'I_3(\gamma)F)^{-1}F'I_3(\gamma)[I_1(\gamma^0)F\beta_1 + (I - I_1(\gamma^0))F\beta_2 + U]$

$\xrightarrow{p} \frac{M_1(\gamma_1^0,\gamma_2)-M_1(\gamma)}{M_3(\gamma)}(\beta_1-\beta_2)+\beta_2;$

$\widehat{\beta}_4(\gamma)=(F'I_4(\gamma)F)^{-1}F'I_4(\gamma)Y$

$=(F'I_4(\gamma)F)^{-1}F'I_4(\gamma)[I_1(\gamma^0)F\beta_1+(I-I_1(\gamma^0))F\beta_2+U]$

$\xrightarrow{p} \frac{M_1(\gamma^0)-M_1(\gamma_1^0,\gamma_2)-M_1(\gamma_1,\gamma_2^0)+M_1(\gamma)}{M_4(\gamma)}(\beta_1-\beta_2)+\beta_2$

and we can get

$\frac{1}{T}RSS_T(\gamma)=\beta_1'M_1(\gamma^0)\beta_1+\beta_2'[Q_{ff}-M_1(\gamma^0)]\beta_2-\sum_{i=1}^4\widehat{\beta}_i'M_i(\gamma)\widehat{\beta}_i+\sigma^2+o_p(1)$

$=\beta_1'M_1(\gamma^0)\beta_1+\beta_2'[Q_{ff}-M_1(\gamma^0)]\beta_2-\sum_{i=1}^4\widehat{\beta}_i'M_i(\gamma)\beta_2-$

$\begin{bmatrix} M_1(\gamma)\widehat{\beta}_1+(M_1(\gamma_1,\gamma_2^0)-M_1(\gamma))\widehat{\beta}_2+[M_1(\gamma_1^0,\gamma_2)-M_1(\gamma)]\widehat{\beta}_3 \\ +[M_1(\gamma^0)-M_1(\gamma_1^0,\gamma_2)-M_1(\gamma_1,\gamma_2^0)+M_1(\gamma)]\widehat{\beta}_4 \end{bmatrix}'(\beta_1-\beta_2)+\sigma^2+$
$o_p(1)$

$=\begin{bmatrix} M_1(\gamma^0)\beta_1-M_1(\gamma)\widehat{\beta}_1-(M_1(\gamma_1,\gamma_2^0)-M_1(\gamma))\widehat{\beta}_2 \\ -[M_1(\gamma_1^0,\gamma_2)-M_1(\gamma)]\widehat{\beta}_3'-[M_1(\gamma^0)-M_1(\gamma_1^0,\gamma_2)-M_1(\gamma_1,\gamma_2^0)+M_1(\gamma)]\widehat{\beta}_4 \end{bmatrix}'(\beta_1-$
$\beta_2)$

$+\sigma^2+o_p(1)$

$=(\beta_1-\beta_2)'\begin{bmatrix} (M_1(\gamma^0)-M_1(\gamma))-\frac{(M_1(\gamma_1,\gamma_2^0)-M_1(\gamma))^2}{M_2(\gamma)}-\frac{(M_1(\gamma_1^0,\gamma_2)-M_1(\gamma))^2}{M_3(\gamma)}- \\ \frac{(M_1(\gamma^0)-M_1(\gamma_1^0,\gamma_2)-M_1(\gamma_1,\gamma_2^0)+M_1(\gamma))^2}{M_4(\gamma)} \end{bmatrix}$

$(\beta_1-\beta_2)+\sigma^2+o_p(1)=(\beta_1-\beta_2)'Q_7(\beta_1-\beta_2)+\sigma^2+o_p(1)$

$=b_1(\gamma)+\sigma^2+o_p(1)$

for any $\gamma_1\le\gamma_1^0$, and $\gamma_2\le\gamma_2^0$, using the following results:

$I_1(\gamma_1,\gamma_2^0)-I_1(\gamma)\le I_2(\gamma),$

$I_1(\gamma_1^0,\gamma_2)-I_1(\gamma)\le I_3(\gamma),$

$I_1(\gamma^0)-I_1(\gamma_1^0,\gamma_2)-I_1(\gamma_1,\gamma_2^0)+I_1(\gamma)\le I_4(\gamma),$

we can prove Q_7 is a positive definite matrix, so $b_1(\gamma)\ge b_1(\gamma^0)=0$ and the equation holds if and only if $\gamma=\gamma^0$.

Case 2: $\gamma_1>\gamma_1^0$, and $\gamma_2<\gamma_2^0$

$I_1(\gamma)I_1(\gamma^0)=I_1(\gamma_1^0,\gamma_2),$

$I_2(\gamma)I_1(\gamma^0)=I_1(\gamma^0)-I_1(\gamma_1^0,\gamma_2),$

$I_3(\gamma)I_1(\gamma^0)=0,$

$I_4(\gamma)I_1(\gamma^0)=0.$

We have

$\widehat{\beta}_1(\gamma) = (F^{'}I_1(\gamma)F)^{-1}F^{'}I_1(\gamma)Y \xrightarrow{p} \frac{M_1(\gamma_1^0,\gamma_2)}{M_1(\gamma)}(\beta_1-\beta_2)+\beta_2;$

$\widehat{\beta}_2(\gamma) = (F^{'}I_2(\gamma)F)^{-1}F^{'}I_2(\gamma)Y \xrightarrow{p} \frac{M_1(\gamma^0)-M_1(\gamma_1^0,\gamma_2)}{M_2(\gamma)}(\beta_1-\beta_2)+\beta_2;$

$\widehat{\beta}_3(\gamma) = (F^{'}I_3(\gamma)F)^{-1}F^{'}I_3(\gamma)Y \xrightarrow{p} \beta_2;$

$\widehat{\beta}_4(\gamma) = (F^{'}I_4(\gamma)F)^{-1}F^{'}I_4(\gamma)Y \xrightarrow{p} \beta_2.$

Then,

$\frac{1}{T}RSS_T(\gamma) = \beta_1^{'}M_1(\gamma^0)\beta_1 + \beta_2^{'}[Q_{ff}-M_1(\gamma^0)]\beta_2 - \sum_{i=1}^{4}\widehat{\beta}_i^{'}M_i(\gamma)\widehat{\beta}_i + \sigma^2 + o_p(1)$
$= \beta_1^{'}M_1(\gamma^0)\beta_1 + \beta_2^{'}[Q_{ff}-M_1(\gamma^0)]\beta_2 - \sum_{i=1}^{4}\widehat{\beta}_i^{'}M_i(\gamma)\beta_2 -$
$\left[M_1(\gamma_1^0,\gamma_2)\widehat{\beta}_1 + [M_1(\gamma^0)-M_1(\gamma_1^0,\gamma_2)]\widehat{\beta}_2\right]^{'}(\beta_1-\beta_2)+\sigma^2+o_p(1)$
$= \left[M_1(\gamma^0)\beta_1 - M_1(\gamma_1^0,\gamma_2)\widehat{\beta}_1^{'} - [M_1(\gamma^0)-M_1(\gamma_1^0,\gamma_2)]\widehat{\beta}_2\right]^{'}(\beta_1-\beta_2)+\sigma^2+o_p(1)$
$= (\beta_1-\beta_2)^{'}\left[M_1(\gamma^0) - \frac{(M_1(\gamma_1^0,\gamma_2))^2}{M_1(\gamma)} - \frac{(M_1(\gamma^0)-M_1(\gamma_1^0,\gamma_2))^2}{M_2(\gamma)}\right](\beta_1-\beta_2)+\sigma^2+$
$o_p(1)$
$= (\beta_1-\beta_2)^{'}Q_8(\beta_1-\beta_2)+\sigma^2+o_p(1)$
$= b_2(\gamma)+\sigma^2+o_p(1).$

for any $\gamma_1 > \gamma_1^0$, and $\gamma_2 < \gamma_2^0$, using the following results:

$I_1(\gamma_1^0,\gamma_2) < I_1(\gamma),$

$I_1(\gamma^0) - I_1(\gamma_1^0,\gamma_2) < I_2(\gamma),$

we can prove that Q_8 is a positive matrix, so $b_2(\gamma) > b_2(\gamma^0) = 0$ for any γ, γ where $\gamma_1 > \gamma_1^0$, and $\gamma_2 < \gamma_2^0$

Since Case 3 and Case 4 are mirror images of Case 2 and Case 1, we can use the similar method to prove that

$\frac{1}{T}RSS_T(\gamma) = \sigma^2 + b_i(\gamma) + o_p(1)$ and $b_i(\gamma) \geq b_i(\gamma^0) = 0$ in both case.

Define a non-stochatic function $b(\gamma)$ as $b_i(\gamma)$ in the ith case, then

$$\sup_{\gamma\in R^2} |\frac{1}{T}RSS_T(\gamma) - b(\gamma) - \sigma^2| = o_p(1)$$

Since $b(\gamma)$ is minimized at $\gamma = \gamma^0$ uniquely, $RSS_T(\gamma)$ is minimized at the true thresholds and $\widehat{\gamma}$ are consistent.

Next, we consider the case where the functional form is misspecified.

The true threshold model is:

$$Y = \sum_{i=1}^{4} I_i(\gamma) F \beta_i + U$$

However, the estimated model is

$$Y = \sum_{i=1}^{4} I_i(\gamma) G \widehat{\beta}_i + \widehat{U}.$$

With a similar projection method used in the paper of Bai et al.(2004), the model can be simply rewritten as a model with new coefficients but without the misspecification problem.

Let $F_t = f(x_t) = (f_1(x_{t1}), ..., f_L(x_{tP}))'$ and similarly, let $G_t = g(x_t)$. Thus

$$F = (F_1, F_2, ..., F_T)',$$

$$G = (G_1, G_2, ..., G_T)'.$$

F is a $T \times P$ matrix and G is a $T \times L$ matrix

Let

$$Q_{fg} = E(F'G).$$

Project F_t on G_t, we have

$$F_t = CG_t + e_t,$$

where C is a $P \times L$ matrix of coefficient. From the orthogonality of G_t and e_t due to the projection, we have, in the limit, $Q_{fg} = CQ_{gg}$.

Let $e = (e_1, e_2, ..., e_T)'$. The model can be rewritten as

$$Y = \sum_{i=1}^{4} I_i \left(GC' + e \right) \beta_i + U = \sum_{i=1}^{4} I_i G \theta_i + W, \tag{35}$$

where

$$\theta_i = C' \beta_i,$$

and

$$W = \sum_{i=1}^{4} I_i e \beta_i + U.$$

Furthermore, from $\beta_{i+1} - \beta_i \neq 0$, we have $\theta_{i+1} - \theta_i = C'(\beta_{i+1} - \beta_i) \neq 0$. The condition has been satisfied by the assumptions.

Thus, the model reduces to that of a standard threshold model with two variables, and G is considered as the true regressors and θ_i as the new coefficients. Now, recall the results from Theorem 1, the estimators are consistent to the true threshold values.

The proof of Theorem 3 is completed by combining the above results.

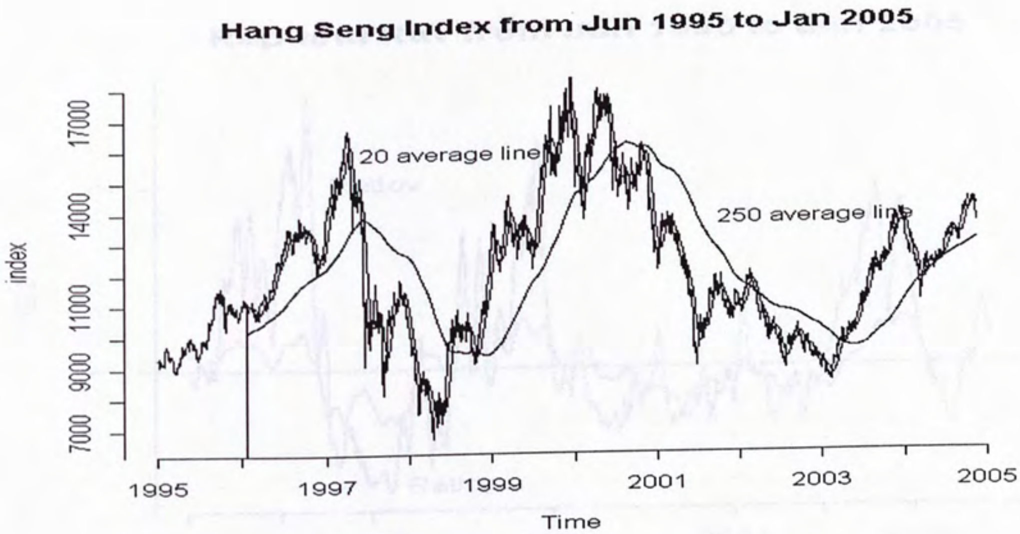

Figure 1: Daily Hang Seng Index and moving average lines.

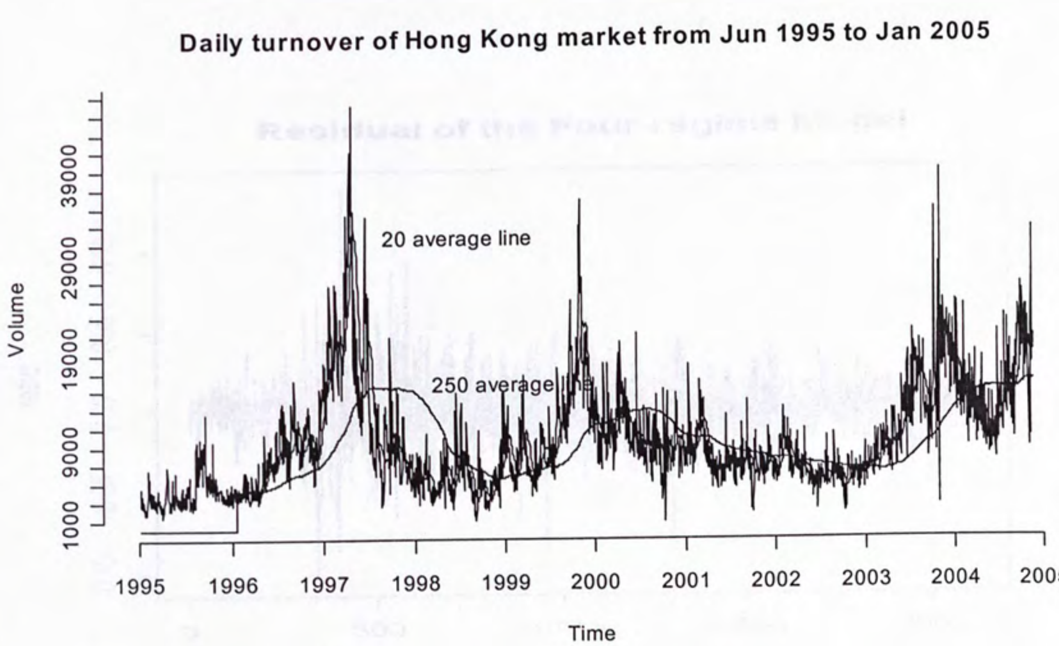

Figure 2: Daily turnover and moving average lines.

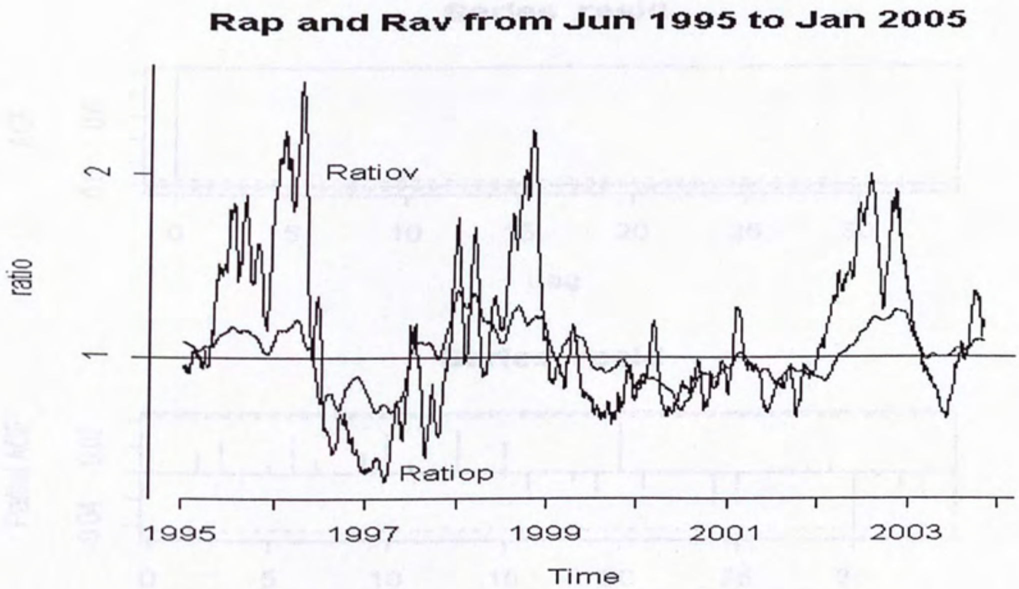

Figure 3: Two threshold variables RAV and RAP.

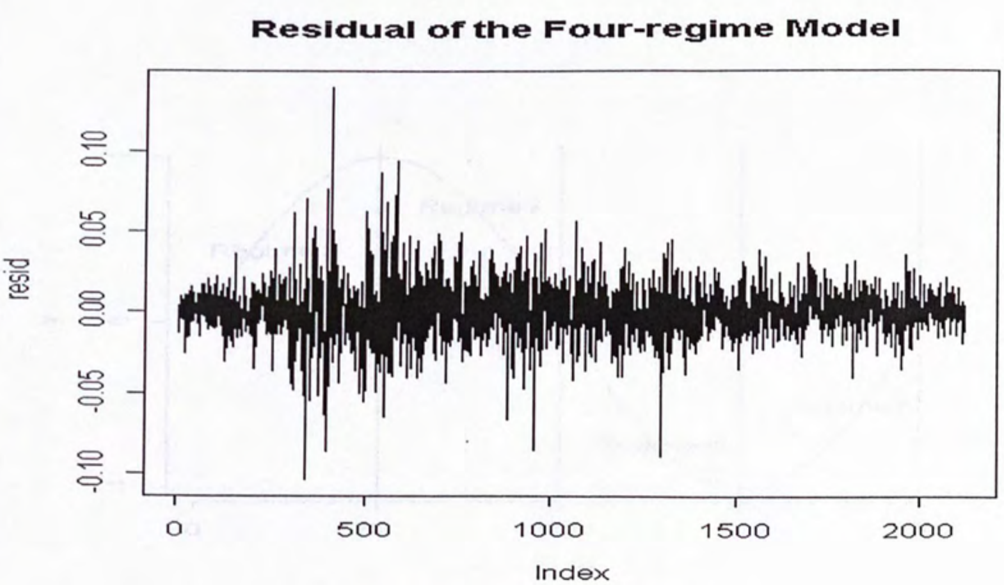

Figure 4: Residual series.

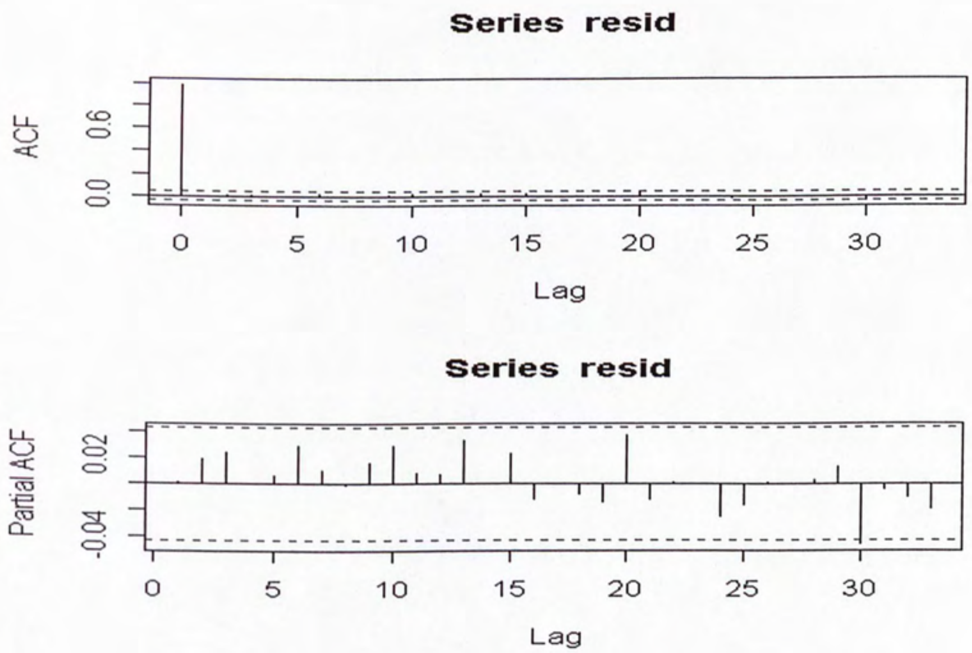

Figure 5: PACF and ACF for residual series

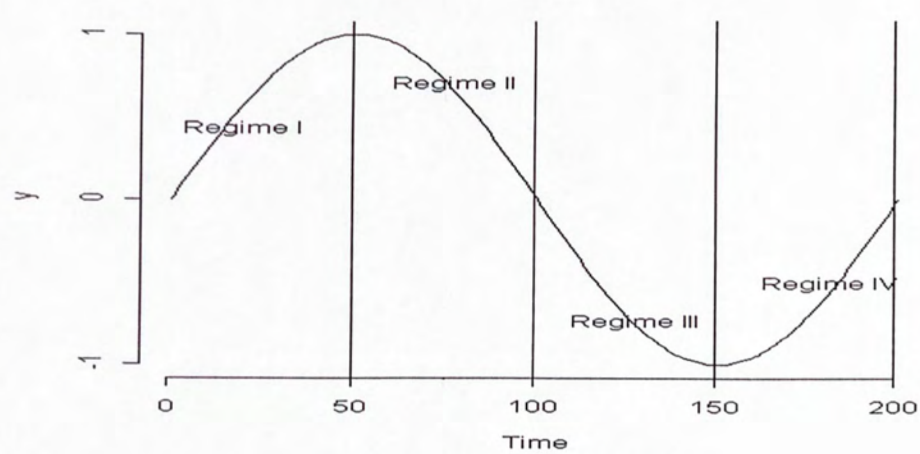

Figure 6: Four-regime cycle.